

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2005, 06:43:25 ; Search time 27 Seconds
(without alignments)
1233.000 Million cell updates/sec

Title: US-10-070-277-2

Perfect score: 1812
Sequence: 1 MEISTQPDMDHILHRDGV.....SGDIIMFAGEMDMLPAPL 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 791:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1467	81.0	377	2 T05124	dhidrooorotase (EC
2	1056.5	58.3	342	2 S76410	dhidrooorotase (EC
3	1039.5	57.4	350	2 E82198	dhidrooorotase VCA
4	983	54.2	344	2 G81934	probable dhidrooor
5	982	54.2	344	2 H81171	dhidrooorotase NMB
6	975	53.8	348	2 T10453	dhidrooorotase (EC
7	971	53.6	348	2 DEECO	dhidrooorotase (EC
8	970	53.5	348	2 D85668	dhidro-oorotase [i
9	970	53.5	348	2 H90808	dhidro-oorotase [i
10	956	52.8	348	2 AF0193	dhidrooorotase (EC
11	951	52.5	393	2 AF2625	dhidrooorotase [im
12	951	52.5	393	2 E97407	dhidrooorotase (dh
13	949	52.4	348	2 AD0638	dhidrooorotase (EC
14	946	52.2	348	1 DEEBOT	dhidrooorotase (EC
15	859.5	47.4	350	2 G84968	dhidrooorotase (EC
16	569	31.4	335	2 B81444	dhidrooorotase (EC
17	566.5	31.3	337	2 T37731	probable dhidrooor
18	541	29.9	339	2 E64592	dhidrooorotase - H
19	527	29.1	339	2 A71922	dhidrooorotase - H
20	481	26.5	364	1 DEBYO	dhidrooorotase (EC
21	466.5	25.7	391	1 DEUSO	dhidrooorotase (EC
22	146	8.1	423	2 A64486	dhidrooorotase (EC
23	140.5	7.8	422	2 C70370	dhidrooorotase - A
24	139	7.7	428	2 B83967	dhidrooorotase pyr
25	131.5	7.3	426	2 A81304	dhidrooorotase hom
26	131	7.2	383	2 T30519	dhidrooorotase (EC
27	131	7.2	430	2 T45151	dhidrooorotase (EC
28	130.5	7.2	426	2 A81676	dhidrooorotase hom
29	129.5	7.1	454	2 A69017	dhidrooorotase - M

ALIGNMENTS

30	128	7.1	424	2 H96963	dhidrooorotase [im
31	120	6.6	441	2 E86758	dhidrooorotase (EC
32	119.5	6.6	422	2 D84403	dhidrooorotase [im
33	116	6.4	439	2 AH2093	dhidrooorotase [im
34	115.5	6.4	417	2 C71212	probable dhidrooor
35	112.5	6.2	404	2 C75027	dhidrooorotase (py
36	112	6.2	427	2 T140167	dhidrooorotase (EC
37	111	6.1	2198	2 T20371	hypothetical prote
38	107.5	5.9	2242	2 A57541	pyrimidine synthe
39	106.5	5.9	422	2 C95135	dhidrooorotase, mu
40	106.5	5.9	428	1 DEBSO	dhidrooorotase (EC
41	106	5.8	376	2 A72391	dhidrooorotase - T
42	105	5.8	359	1 D70920	probable moa prot
43	104.5	5.8	424	2 D89892	dhidrooorotase [im
44	103.5	5.7	416	2 E75437	dhidrooorotase - D
45	103.5	5.7	422	2 E98003	dhidrooorotase (EC

RESULT 1

T05124
dhidrooorotase (EC 3.5.2.3) - Arabidopsis thaliana

N,Alternative names: protein F7H19.110

C,Species: Arabidopsis thaliana (mouse-ear cress)

C,Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004

C,Accession: T05124

R,Bevan, M.; Peters, S.A.; Van Staeren, M.; Dirke, W.; Stiekema, W.; Bancroft, I.; Me

submitted to the Protein Sequence Database, July 1998

A,Reference number: Z15399

A,Accession: T05124

A,Molecule type: DNA

A,Residues: 1-377 <BEV>

A,Cross-references: UNIPROT:O04904; EMBL:AL031018

A,Experimental source: cultivar Columbia; BAC clone F7H19

C,Genetics:

A,Map position: 4

A,Intons: 10/3; 60/2; 121/1; 166/3; 222/1; 261/1; 314/3

A,Note: F7H19.110

C,Function:

A,Description: catalyzes conversion of N-carbamoyl-L-aspartate into dhidrooorotase

C,Keywords: hydrolase; pyrimidine nucleotide biosynthesis; zinc

Query Match 81.0%; Score 1467; DB 2; Length 377;
Best Local Similarity 78.4%; Pred. No. 1.5e-113;

Matches 269; Conservative 41; Mismatches 33; Indels 0; Gaps 0;

QY	1	MEISTQPDMDHILHRDGV	LKAVVSHSAHFGRAIMPTLTTTAAVAAREALIK	60
DB	33	MEITITQPDMDHILHRDGL	HLVPHSASNFRAIYMPILKEPVSTAAAIYRKIMK	92
QY	61	SLPVDSENFIMLTLYLTDT	SPMEIKLARSQVFGVLYPAAGATTSQDVTDLFGKCL	120
DB	93	ALPSESSFDPLMTLYLTDL	KPLPEIRKLARSGVVAVKLPAGATTSQDVTDLFGKCL	152
QY	121	PVLQEWENHMPPLVHGEV	TNPPEVMDREKVFIEYVLRPLVQKFPOLKVMMEHTTIDA	180
DB	153	PVLEEMVKQMPPLVHGEV	TDPSIDVDFREKIFIEYVLRPLVQKFPOLKVMMEHTTIDA	212
QY	181	VKPYESTGEGVAAATTP	POHLVLRNSLPQGGLOPHNYCPLVYKREIHRALVAATVSGS	240
DB	213	VNFVESCSEGSVGVATVP	POHLVLRNSLPQGGLOPHNYCPLVYKREIHRALVAATVSGS	272
QY	241	KRPFLEGTDSA	PHDRRREKSGCAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF	300
DB	273	KKPFLEGTDSA	PHDRRREKSGCAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF	332
QY	301	YGLPRNNSKI	KLSTKTPKVPESFYSAGDIIIMFAGEMDMLP	343
DB	333	YGLPRNNSKIT	LTKKSPKVPDFVNFPGELIIMPAGETLQWOP	375

Db 122 LEBMAKQGLFLVHGVTDPEDIDIFREAAFIKRWKPKVLQVQPNLKVPEHITTEAAR 181
Qy 183 FVESCIEGFVAATVTPQHLVLNNSLFOGGLQPHNYCLPVLKREIHREALVSAVTSKSK- 241
Db 182 LVLEAGDN-VAAVTVTQHLNLRNDLLVGVSRPHHCLPLKRETHQRLVAAVT-GEKA 239
Qy 242 -RFFLTGDSAPHDRRRKCCSCGAGIYNAPVALSVYAKVFEKNALDKLEAFISFNGPDF 300
Db 240 HKFPLGTDSPAHKSAKENACGAGMFSAMTAILEYAEVFEKAGALDKLEAFASKGAPR 299
Qy 301 YGLPRNNSKIKLSKTPMKVPESPSYASGDIIPFAGEMLDW 341
Db 300 YGIPENTDTITLVKQSQVTPASIPYGDDELVPWRAGSEIGW 340

RESULT 5
H81171
dihydroorotase NMB0682 [imported] - *Neisseria meningitidis* (strain MC58 serogroup B)
C/Species: *Neisseria meningitidis*
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C/Accession: H81171
R/Retelid, H.; Saudere, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Hatt, D.H.; Salzman, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
rt, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizze, M.
Science 287, 1809-1815, 2000
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A/Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A/Reference number: A81000; MUID:20175755; PMID:10710307
A/Accession: H81171
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-344 <STO>
A/Cross-references: UNIPROT:Q9K0D1; GB:AE002422; GB:AE002098; NID:G7225898; PIDN:AAE4110
A/Experimental source: serogroup B, strain MC58
C/Genetics:
A/Gene: NMB0682
C/Superfamily: dihydroorotase

Query Match 54.2%; Score 982; DB 2; Length 344;
Best Local Similarity 56.3%; Pred. No. 1.7e-73;
Matches 192; Conserved 55; Mismatches 88; Indels 6; Gaps 5;

Qy 3 LSTIPDDMHLNRDGVILKAVVSHSAHFGRAIWMPLKPTTTAAVAAYREALIKSL 62
Db 4 LTIIRPDMDHLNRDGVILKAVVSHSAHFGRAIWMPLKPTTTAAVAAYREALIKSL 63
Qy 63 PVSDENPLMTLYLTDTTSPMEIKLARESOVFGVKLYPAGATTNSODGVTDLFGKCLPV 122
Db 64 PEGSAFEPMLTYLTDTTSPMEIKLARESOVFGVKLYPAGATTNSODGVTDLFGKCLPV 121
Qy 123 LOEVENHNMPLVHGEVTNPEVDFREKVFIEVLRPLVQKFPOLKVMEHVTITDAVK 182
Db 122 LEBMAKQGLFLVHGVTDPEDIDIFREAAFIKRWKPKVLQVQPNLKVPEHITTEAAR 181
Qy 183 FVESCIEGFVAATVTPQHLVLNNSLFOGGLQPHNYCLPVLKREIHREALVSAVTSKSK- 241
Db 182 LVLEAGDN-VAAVTVTQHLNLRNDLLVGVSRPHHCLPLKRETHQRLVAAVT-GEKA 239
Qy 242 -RFFLTGDSAPHDRRRKCCSCGAGIYNAPVALSVYAKVFEKNALDKLEAFISFNGPDF 300
Db 240 HKFPLGTDSPAHKSAKENACGAGMFSAMTAILEYAEVFEKAGALDKLEAFASKGAPR 299
Qy 301 YGLPRNNSKIKLSKTPMKVPESPSYASGDIIPFAGEMLDW 341
Db 300 YGIPENTDTITLVKQSQVTPASIPYGDDELVPWRAGSEIGW 340

RESULT 6
T10453
dihydroorotase (EC 3.5.2.3) - *Pseudomonas aeruginosa*
C/Species: *Pseudomonas aeruginosa*
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C/Accession: T10453; E83204

R.Bricta, D.M.; Brown, T.M.; Houghton, J.E.; O'Donovan, G.A.
submitted to the EMBL Data Library, November 1998
A/Description: *Pyrc* (dihydroorotase) sequence from *P. aeruginosa*.
A/Reference number: Z17024
A/Accession: T10453
A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA
A/Residues: 1-348

A/Cross-references: UNIPROT:P72170; EMBL:U73505; NID:G3868711; PID:G3868712
R/Stover, C.K.; Pham, X.Q.; Etwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lm,
-; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic path
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: E83204
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-348 <STO>
A/Cross-references: GB:AE004773; GB:AE004091; NID:G9949672; PIDN:AA06915.1; GSPDB:GN001
A/Experimental source: strain PA01
C/Genetics:
A/Gene: *pyrc*; PA3527
C/Function:
A/Description: catalyzes conversion of N-carbamoyl-L-aspartate into dihydroorotase
A/Pathway: pyrimidine nucleotide biosynthesis
C/Superfamily: dihydroorotase
C/Keywords: hydrolase; pyrimidine nucleotide biosynthesis; zinc

Query Match 53.8%; Score 975; DB 2; Length 348;
Best Local Similarity 55.5%; Pred. No. 6.4e-73;
Matches 188; Conserved 43; Mismatches 106; Indels 2; Gaps 2;

Qy 3 LSTIPDDMHLNRDGVILKAVVSHSAHFGRAIWMPLKPTTTAAVAAYREALIKSL 62
Db 5 LTIIRPDMDHLNRDGVILKAVVSHSAHFGRAIWMPLKPTTTAAVAAYREALIKSL 64
Qy 63 PVSDENPLMTLYLTDTTSPMEIKLARESOVFGVKLYPAGATTNSODGVTDLFGKCLPV 122
Db 65 PAASRFEPLMTLYLTDTTSPMEIKLARESOVFGVKLYPAGATTNSODGVTDLFGKCLPV 123
Qy 123 LOEVENHNMPLVHGEVTNPEVDFREKVFIEVLRPLVQKFPOLKVMEHVTITDAVK 182
Db 124 LEBMAKQGLFLVHGVTDPEDIDIFREAAFIKRWKPKVLQVQPNLKVPEHITTEAAR 183
Qy 183 FVESCIEGFVAATVTPQHLVLNNSLFOGGLQPHNYCLPVLKREIHREALVSAVTSKSK- 242
Db 184 LVLEAGDN-VAAVTVTQHLNLRNDLLVGVSRPHHCLPLKRETHQRLVAAVT-GEKA 242
Qy 243 FFLGTGDSAPHDRRRKCCSCGAGIYNAPVALSVYAKVFEKNALDKLEAFISFNGPDFY 302
Db 243 FFLGTGDSAPHDRRRKCCSCGAGIYNAPVALSVYAKVFEKNALDKLEAFISFNGPDFY 302
Qy 303 LPRNNSKIKLSKTPMKVPESPSYASGDIIPFAGEMLDW 341
Db 303 LPRNNSKIKLSKTPMKVPESPSYASGDIIPFAGEMLDW 341

RESULT 7
D8ECO
dihydroorotase (EC 3.5.2.3) - *Escherichia coli* (strain K-12)
N/Alternate names: carbamoylasepartic dehydrase
C/Species: *Escherichia coli*
C/Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 09-Jul-2004
C/Accession: A25008; A27084; C64849
R/Backstrom, D.; Sjoberg, R.M.; Lundberg, L.G.
Eur. J. Biochem. 160, 77-82, 1986
A/Title: Nucleotide sequence of the structural gene for dihydroorotase of *Escherichia c*
A/Reference number: A91177; MUID:87030260; PMID:2876892
A/Accession: A25008
A/Molecule type: DNA
A/Residues: 1-348 <BAC>
A/Cross-references: UNIPROT:P05020; EMBL:X04469; NID:G42605; PIDN:CAA28157.1; PID:G4260

A:Experimental source: strain K-12
R:Wilson, H.R.; Chan, P.T.; Turnbough Jr., C.L.
J. Bacteriol. 169: 3051-3058, 1987
A:Title: Nucleotide sequence and expression of the *pyrC* gene of *Escherichia coli* K-12.
A:Reference number: A91836; MUID:87250268; PMID:2885307
A:Accession: A27084
A:Molecule type: mRNA
A:Residues: 1-348 <WIL>
A:Cross-references: EMBL:M6752; NID:g147472; PIDN:AAA24482.1; PID:g147473
A:Experimental source: strain K-12
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64849
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-348 <BLAT>
A:Experimental source: GB:AE000207; GB:U00096; NID:g1787293; PIDN:AAC74146.1; PID:g1787301,
C:Comment: The de novo synthesis of UMP, the precursor of all pyrimidine nucleotides, is
reased primarily by a cytidine nucleotide.
C:Genetics:
A:Gene: *pyrC*
A:Map position: 23 min
C:Function:
A:Description: catalyzes conversion of N-carbamoyl-L-aspartate into dihydroorotate
A:Pathway: pyrimidine nucleotide biosynthesis
A:Note: zinc required for catalytic activity
C:Superfamily: dihydroorotase
C:Keywords: homodimer; hydrolase; pyrimidine nucleotide biosynthesis; zinc
F:17.19/Binding site: zinc (his) #status predicted

Query Match 53.6%; Score 971; DB 1; Length 348;
Best Local Similarity 55.8%; Pred. No. 1.4e-72;
Matches 189; Conservative 46; Mismatches 102; Indels 2; Gaps 2;

QY 3 LSTGPDMMHLNRDGVKAVVSHAHFGRATVWNLKPPITTTAAVAAYREAILKSL 62
DB 8 LKIRPDDMMHLNRDGMKTVVPTSEIYGRATVWNLKPPITTTAAVAAYREAILHAY 67
QY 63 PVDSDPNLMTLYLDTTSPMEIKLARESQVVGKLYPAGATTNSODGVTDFGKCLPV 122
DB 68 PAGHDFPLMTCYLTDLDPELEKRGNEGVFTAAKLYPAAATTNSHGVTSV-DAIMPV 126
QY 123 LOEWVHNMPPLVHGEVTNPEDVMEFDEKVFIEFTVLRPLVOKFPOLKVMHEVTTIDAVK 182
DB 127 LERMEKIGWPLVHGEVTHADIDIFREARFISVMEPLRQRLTALKVFEHITTKDAD 186
QY 183 FVESCIEGVAAATVTPQHLVLRNSLFOGGLQPHNYCLPVLRKRIHREALVSAVTSQSKR 242
DB 187 YVRDGNR-LAAITTPQHLMFNNHMLVGVGRPHLYCLDLKRNTHQQLRELVAAGFNK 245
QY 243 FFLGDSAPHRDRRKECCGACAGIYVAPVALSVYAKFEKENALDKLEAFTSFGSDPYG 302
DB 246 VFLGDSAPHRDRRKECCGACAGCFNAPVALSVYAKFEKENALDHFEAFCSVNGPQYFG 305
QY 303 LPRNNSKIKLSKTPWKPESPSYASGDIIPMPGEMLDW 341
DB 306 LPVNDFTIELVREHQVAVSIALTDTLVLPFLAGETVRW 344

RESULT 8
D85668
dihydro-orotase [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)
C:Species: *Escherichia coli*
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: D85668
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, R.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: D85668
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-348 <STO>
A:Cross-references: UNIPROT:Q8X8N8; GB:AE005174; NID:g12514597; PIDN:AAG55808.1; GSPDB
C:Experimental source: strain O157:H7, substrain EDL933
A:Gene: *pyrC*
C:Superfamily: dihydroorotase

Query Match 53.5%; Score 970; DB 2; Length 348;
Best Local Similarity 55.8%; Pred. No. 1.7e-72;
Matches 189; Conservative 46; Mismatches 102; Indels 2; Gaps 2;

QY 3 LSTGPDMMHLNRDGVKAVVSHAHFGRATVWNLKPPITTTAAVAAYREAILKSL 62
DB 8 LKIRPDDMMHLNRDGMKTVVPTSEIYGRATVWNLKPPITTTAAVAAYREAILHAY 67
QY 63 PVDSDPNLMTLYLDTTSPMEIKLARESQVVGKLYPAGATTNSODGVTDFGKCLPV 122
DB 68 PAGHDFPLMTCYLTDLDPELEKRGNEGVFTAAKLYPAAATTNSHGVTSV-DAIMPV 126
QY 123 LOEWVHNMPPLVHGEVTNPEDVMEFDEKVFIEFTVLRPLVOKFPOLKVMHEVTTIDAVK 182
DB 127 LERMEKIGWPLVHGEVTHADIDIFREARFISVMEPLRQRLTALKVFEHITTKDAD 186
QY 183 FVESCIEGVAAATVTPQHLVLRNSLFOGGLQPHNYCLPVLRKRIHREALVSAVTSQSKR 242
DB 187 YVRDGNR-LAAITTPQHLMFNNHMLVGVGRPHLYCLDLKRNTHQQLRELVAAGFNK 245
QY 243 FFLGDSAPHRDRRKECCGACAGIYVAPVALSVYAKFEKENALDKLEAFTSFGSDPYG 302
DB 246 VFLGDSAPHRDRRKECCGACAGCFNAPVALSVYAKFEKENALDHFEAFCSVNGPQYFG 305
QY 303 LPRNNSKIKLSKTPWKPESPSYASGDIIPMPGEMLDW 341
DB 306 LPVNDFTIELVREHQVAVSIALTDTLVLPFLAGETVRW 344

RESULT 9
H90808
dihydro-orotase [imported] - *Escherichia coli* (strain O157:H7, substrain RMD 0509952)
C:Species: *Escherichia coli*
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: H90808
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Iehi, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: H90808
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-348 <HAY>
A:Cross-references: UNIPROT:Q8X8N8; GB:BA000007; PIDN:BA834863.1; PID:g13360904; GSPDB:C
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: *pyrC*
C:Superfamily: dihydroorotase

Query Match 53.5%; Score 970; DB 2; Length 348;
Best Local Similarity 55.8%; Pred. No. 1.7e-72;
Matches 189; Conservative 46; Mismatches 102; Indels 2; Gaps 2;

QY 3 LSTGPDMMHLNRDGVKAVVSHAHFGRATVWNLKPPITTTAAVAAYREAILKSL 62
DB 8 LKIRPDDMMHLNRDGMKTVVPTSEIYGRATVWNLKPPITTTAAVAAYREAILHAY 67
QY 63 PVDSDPNLMTLYLDTTSPMEIKLARESQVVGKLYPAGATTNSODGVTDFGKCLPV 122
DB 68 PAGHDFPLMTCYLTDLDPELEKRGNEGVFTAAKLYPAAATTNSHGVTSV-DAIMPV 126

QY 3 LSTGPDMDHLHLDGDLVKAUVSHSAHREGRAIWMNKPIPTTTAAVAAREALIKSL 62
 Db 52 LTLRRPDDMDHLHLDGDLVKAUVSHSAHREGRAIWMNKPIPTTTAAVAAREALIKSL 111
 QY 63 PVSDENPMLTYLTDTTSPMEIKLARESOVGVKLYPAAGATTSGQDVTDLFGKCLPV 122
 Db 112 PEGDRFEPMLTYLTDTTSPMEIKLARESOVGVKLYPAAGATTSGQDVTDLFGKCLPV 170
 QY 123 LOEMVHNHMLLVHGEVNTPEVDMFDEKVFETVLRPLVQKPOLKVMMEHVTITDAVK 182
 Db 171 LERNAKGLPLCHGEVNTPEVDMFDEKVFETVLRPLVQKPOLKVMMEHVTITDAVK 230
 QY 183 FVESCTEGFVAATVTPQHLVNLNRSIFOGGLOPHNYCPLVLRREIHRALVSAVTSQSKR 242
 Db 231 YIKS-SNANLAGSITTHHLIINNALLVGIRPHYCLPVAKREHRLALRAATSGDAR 289
 QY 243 FFLGTSAPHRDRRKECGCGAGIYNAPALSVYAKVEKENALDKLEAFTSFNGPDYFG 302
 Db 290 FFLGTSAPHRDRRKECGCGAGIYNAPALSVYAKVEKENALDKLEAFTSFNGPDYFG 349
 QY 303 LPRNNSKIKLSTKPMKRV--PESFSYASGDI--PMPAGEMLDW 341
 Db 350 LPRNNSKIKLSTKPMKRV--PESFSYASGDI--PMPAGEMLDW 390

RESULT 13

AD0638
 dihydrooocatae (EC 3.5.2.3) [imported] - *Salmonella enterica* subsp. *enterica* serovar Typh
 C/Species: *Salmonella enterica* subsp. *enterica* serovar Typh
 A/Note: This sequence has also been called *Salmonella typh*
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C/Accession: AD0638
 R/Parthill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 Th. T.; Connerton, P.; Cronin, A.; Davys, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A/Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
 A/Reference number: AB0502; MUID:21534947; PMID:11675608
 A/Accession: AD0638
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-348 <PAR>
 A/Cross-references: GB:AL513382; PIDN:CAD08287.1; PID:g16502333; GSPDB:GN00176
 A/Gene: STY1201
 C/Superfamily: dihydrooocatae
 C/Keywords: hydrolase

Query Match 52.4%; Score 949; DB 2; Length 348;
 Best Local Similarity 54.3%; Pred. No. 9e-71;
 Matches 184; Conservative 49; Mismatches 104; Indels 2; Gaps 2;

QY 3 LSTGPDMDHLHLDGDLVKAUVSHSAHREGRAIWMNKPIPTTTAAVAAREALIKSL 62
 Db 8 LKIRPDDMDHLHLDGDLVKAUVSHSAHREGRAIWMNKPIPTTTAAVAAREALIKSL 67
 QY 63 PVSDENPMLTYLTDTTSPMEIKLARESOVGVKLYPAAGATTSGQDVTDLFGKCLPV 122
 Db 68 PACHDFTPLMTCTYLTSLDADELERGFHGVFTAAKLIPANAKTNSHGVTSV-DALMPV 126
 QY 123 LOEMVHNHMLLVHGEVNTPEVDMFDEKVFETVLRPLVQKPOLKVMMEHVTITDAVK 182
 Db 127 LERNAKGLPLCHGEVNTPEVDMFDEKVFETVLRPLVQKPOLKVMMEHVTITDAVK 186
 QY 183 FVESCTEGFVAATVTPQHLVNLNRSIFOGGLOPHNYCPLVLRREIHRALVSAVTSQSKR 242
 Db 187 YIKS-SNANLAGSITTHHLIINNALLVGIRPHYCLPVAKREHRLALRAATSGDAR 245
 QY 243 FFLGTSAPHRDRRKECGCGAGIYNAPALSVYAKVEKENALDKLEAFTSFNGPDYFG 302
 Db 246 APLGTSAPHRDRRKECGCGAGIYNAPALSVYAKVEKENALDKLEAFTSFNGPDYFG 305

QY 303 LPRNNSKIKLSTKPMKRV--PESFSYASGDI--PMPAGEMLDW 341
 Db 306 LPRNNSKIKLSTKPMKRV--PESFSYASGDI--PMPAGEMLDW 344

RESULT 14

DEEBOT
 dihydrooocatae (EC 3.5.2.3) - *Salmonella typhimurium*
 C/Species: *Salmonella typhimurium*
 C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
 C/Accession: A27143
 R/Neuhard, U.; Kellin, R.A.; Stauning, E.
 Eur. J. Biochem. 157, 335-342, 1986
 A/Title: Cloning and structural characterization of the *Salmonella typhimurium* *pyrC* gene
 A/Reference number: A27143; MUID:86220211; PMID:2872051
 A/Accession: A27143
 A/Molecule type: DNA
 A/Residues: 1-348 <NEU>
 A/Cross-references: UNIPROT:P06204; GB:X03928; NID:947865; PIDN:CAA27567.1; PID:947867
 A/Experimental source: strain Lr2
 C/Comment: The de novo synthesis of UMP, the precursor of all pyrimidine nucleotides, is
 repressed primarily by a cytidine nucleotide.
 A/Gene: *pyrC*
 A/Map position: 23 min
 C/Function:
 A/Description: catalyzes conversion of N-carbamoyl-L-aspartate into dihydrooocatae
 A/Pathway: pyrimidine nucleotide biosynthesis
 A/Note: zinc required for catalytic activity
 C/Superfamily: dihydrooocatae
 C/Keywords: homodimer; hydrolase; pyrimidine nucleotide biosynthesis; zinc
 F17.19/binding site: zinc (His) #status predicted

Query Match 52.2%; Score 946; DB 1; Length 348;
 Best Local Similarity 54.0%; Pred. No. 1.6e-70;
 Matches 183; Conservative 50; Mismatches 104; Indels 2; Gaps 2;

QY 3 LSTGPDMDHLHLDGDLVKAUVSHSAHREGRAIWMNKPIPTTTAAVAAREALIKSL 62
 Db 8 LKIRPDDMDHLHLDGDLVKAUVSHSAHREGRAIWMNKPIPTTTAAVAAREALIKSL 67
 QY 63 PVSDENPMLTYLTDTTSPMEIKLARESOVGVKLYPAAGATTSGQDVTDLFGKCLPV 122
 Db 68 PACHDFTPLMTCTYLTSLDADELERGFHGVFTAAKLIPANAKTNSHGVTSV-DALMPV 126
 QY 123 LOEMVHNHMLLVHGEVNTPEVDMFDEKVFETVLRPLVQKPOLKVMMEHVTITDAVK 182
 Db 127 LERNAKGLPLCHGEVNTPEVDMFDEKVFETVLRPLVQKPOLKVMMEHVTITDAVK 186
 QY 183 FVESCTEGFVAATVTPQHLVNLNRSIFOGGLOPHNYCPLVLRREIHRALVSAVTSQSKR 242
 Db 187 YIKS-SNANLAGSITTHHLIINNALLVGIRPHYCLPVAKREHRLALRAATSGDAR 245
 QY 243 FFLGTSAPHRDRRKECGCGAGIYNAPALSVYAKVEKENALDKLEAFTSFNGPDYFG 302
 Db 246 APLGTSAPHRDRRKECGCGAGIYNAPALSVYAKVEKENALDKLEAFTSFNGPDYFG 305
 QY 303 LPRNNSKIKLSTKPMKRV--PESFSYASGDI--PMPAGEMLDW 341
 Db 306 LPRNNSKIKLSTKPMKRV--PESFSYASGDI--PMPAGEMLDW 344

RESULT 15

G84968
 dihydrooocatae (EC 3.5.2.3) [imported] - *Buchnera* sp. (strain APS)
 C/Species: *Buchnera* sp.
 C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C/Accession: G84968
 R/Chisembu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
 Nature 407, 81-86, 2000
 A/Title: Genome sequence of the endosymbiotic bacterial symbiont of aphids *Buchnera* sp. *j*
 A/Reference number: AB4930; MUID:20445173; PMID:10993077

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2005, 01:28:38 ; Search time 5666 Seconds
(without alignments)
10869.499 Million cell updates/sec

Title: US-10-070-277-1

Perfect score: 1271

Sequence: 1 ttgcaaaaatgagctctca.....tgaagattcactgataaaaa 1271

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: 1: gb_Da: 2: gb_Htg: 3: gb_In: 4: gb_Om: 5: gb_Ov: 6: gb_Pat: 7: gb_Pn: 8: gb_Pl: 9: gb_Pr: 10: gb_Ro: 11: gb_Sts: 12: gb_Sy: 13: gb_Un: 14: gb_Vi:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1271	100.0	1271	6	AX088927	Sequence
2	1271	100.0	1271	6	AX093580	Sequence
3	646	50.8	1743	6	AX470409	Sequence
4	643.2	50.6	1038	8	BVU251897	Beta vulg
5	587	46.2	1362	8	ATFA000146	Arabidops
6	549.2	43.2	1750	8	AK122010	Oryza sat
7	345.6	27.2	130001	1	D90915	Synechocyst
8	294.2	23.1	306050	1	BX321858	Nitrosomo
9	290.6	22.9	303450	1	AP005085	Beta vulg
10	282.6	22.2	208524	1	AE016925	Chromobac
11	281	22.1	1071	6	AR375107	Sequence
12	281	22.1	11619	1	AE004420	Vibrio ch
13	280.4	22.1	349082	1	BX572091	Prochloro
14	277.2	21.8	10421	1	AE013777	Yersinia
15	277.2	21.8	110000	1	BX936398	Continuatio
16	277.2	21.8	193050	1	AE014149	Yersinia
17	277.2	21.8	291326	1	AE017135	Yersinia
18	276.4	21.7	348672	1	CR378670	Photobact
19	275.2	21.7	190050	1	AL646059	Ralstonia

20	274.6	21.6	110000	1	AE017282_22	Continuation (23 o
21	273	21.5	253750	1	AP005345	AP005345 Vibrio vu
22	272.8	21.5	182624	6	AX770901	Sequence
23	272.8	21.5	348613	1	BX571865	Photobact
24	269.8	21.2	336182	1	AE016813	Vibrio vu
25	267.4	21.0	1041	6	AR318865	Sequence
26	265.8	20.9	300355	1	AE017162	Prochloro
27	265.2	20.9	110000	1	BX571965	Continuation (35 o
28	265.2	20.9	110000	1	CP000010_25	Continuation (26 o
29	264.2	20.8	110000	6	AR409405_3	Continuation (4 of
30	264.2	20.8	110000	6	BD061520_3	Continuation (4 of
31	264.2	20.8	293181	1	AP001119	AP001119 Buchera
32	263.6	20.7	349213	1	BX569693	Synechoc
33	262.8	20.7	110000	1	BX950851_20	Continuation (21 o
34	261.4	20.6	349323	1	BX572096	Prochloro
35	260.6	20.5	10909	1	AE007977	Agrobacte
36	260.6	20.5	13047	1	AE009010	Agrobacte
37	257	20.2	10986	1	AE015804	Shewanell
38	256.2	20.2	11088	1	AE014109	Buchera
39	254.8	20.0	300490	1	AE016965	Coxiella
40	245.6	19.3	1629	1	ECOPYRCA	M16752 E. coli K-12
41	245.6	19.3	2046	1	ECOPYRCA	X04469 E. coli pyr
42	245.6	19.3	12029	1	AE005316	Escherich
43	245.6	19.3	19201	1	D90742	Escherichia
44	245.6	19.3	19418	1	D90743	Escherichia
45	245.6	19.3	110000	1	U00096_11	Continuation (12 o

ALIGNMENTS

RESULT 1	AX088927	1271 bp	DNA	linear	PAT 17-MAR-2001
LOCUS	AX088927	Sequence 1 from Patent WO0114569.			
DEFINITION	AX088927				
ACCESSION	AX088927				
VERSION	AX088927.1	GI:13397685			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
1	Enhardt, T., Stitt, M., Geigenberger, P. L., Loeff, I., Ziemer, R. and Schroeder, M.				
TITLE					
JOURNAL					
1	Increasing the polysaccharide content in plants				
2	Patent: WO 014569-A1 01-MAR-2001,				
3	Basf Aktiengesellschaft (DE)				
4	Location/Qualifiers				
5	1. 1271				
6	/organism="Solanum tuberosum"				
7	/mol_type="unassigned DNA"				
8	/db_xref="taxon:4113"				
9	_1049				
10	/note="unassigned protein product"				
11	/codon_start=1				
12	/protein_id="CAC34708.1"				
13	/db_xref="GI:13397686"				
14	/translation="MELSDTPDDMLHLRDGDLKAVSHSAHFGRAIYVNLKPP				
15	ITTTAAVAAYREAIKSLPDSDEPNPLMTLYLTPTSPMEIKLAREQVGVGLYPA				
16	GATNSODGVTLDLPGKCLPYLQEVENHMLLVHGEVTNPEVDDEKVEIEVLRP				
17	LVQKFPOLKVVMEHTTIDAVKRFVESTCEGVAATVYPOHVINRSLFOGGLPHNY				
18	CLPYLKEIHREALVSAVTSKSKFFIGTDSAPDRRKECCSCCAITNAPVALSVY				
19	AKYFEKENALDKLAFTSPNGPDYGLPRNNSKIKIKTLPKVPESFVASGDIIPMF				
20	AGEMLDWLPAPL"				

ORIGIN

Query Match 100.0%; Score 1271; DB 6; Length 1271;
Best local Similarity 100.0%; Pred. No. 4.8e-313; Indels 0; Gaps 0;
Matches 1271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 1 TTGCAAAAATGAGCTCTCATCAACAACTGATGATTGACATCTTCATCTCGTGATG 60

Db	1	TTGCAGAAAAATGAGAGCTTCATACACAACTGATGATGGCATCTTATCTCCGATG	60
Qy	61	GTGATGTTCTTAAGGAGAGTTGTCTCTCAAGTGCATATCACTTTGGAGGGCAATATGTCA	12
Db	61	GTGATGTTCTTAAGGAGAGTTGTCTCTCAAGTGCATATCACTTTGGAGGGCAATATGTCA	12
Qy	121	TGCCAAATTTGAGAGCTCTCATCACTGCTGCTGTGATACATACCGGAGGGCGCA	18
Db	121	TGCCAAATTTGAGAGCTCTCATCACTGCTGCTGTGATACATACCGGAGGGCGCA	18
Qy	181	TATTGAAATCTTTACTCTGTGATAGTGAATTTCAACCTCTTATGACACTTTATTTGACAG	24
Db	181	TATTGAAATCTTTACTCTGTGATAGTGAATTTCAACCTCTTATGACACTTTATTTGACAG	24
Qy	241	ATACACACAGTCTTAATGAGAAATCAAACTTAGACAAGAGAGACAGGTCGTATTTGGGGTGA	30
Db	241	ATACACACAGTCTTAATGAGAAATCAAACTTAGACAAGAGAGACAGGTCGTATTTGGGGTGA	30
Qy	301	AGTTGTAACCGTGTGTCGCCACGCAAAATTTCTCAAGATGAGTGAATCTTTTCGGGA	36
Db	301	AGTTGTAACCGTGTGTCGCCACGCAAAATTTCTCAAGATGAGTGAATCTTTTCGGGA	36
Qy	361	AGTGTTTACAGATTCACAAAGAAATGAGTGAATTAATGAGTCTGTGTTATGAGAG	42
Db	361	AGTGTTTACAGATTCACAAAGAAATGAGTGAATTAATGAGTCTGTGTTATGAGAG	42
Qy	421	AGGTTACTAATCTGAGGTGACATGTTTGAATAGAGAAAGATTAATGATGAAACGGTTC	48
Db	421	AGGTTACTAATCTGAGGTGACATGTTTGAATAGAGAAAGATTAATGATGAAACGGTTC	48
Qy	481	TAAAGACCGTGTGTCACAAATTTCCAAATTTGAAGGTCGTGATGAGACATGTTTACCA	54
Db	481	TAAAGACCGTGTGTCACAAATTTCCAAATTTGAAGGTCGTGATGAGACATGTTTACCA	54
Qy	541	TTGATGCTGTTAAGTTGTTGAATCTTGCACGTGAAGGATTTGTTGACGCAACTGTACCC	60
Db	541	TTGATGCTGTTAAGTTGTTGAATCTTGCACGTGAAGGATTTGTTGACGCAACTGTACCC	60
Qy	601	CACAACATCTGTTTGTGAAACAGAAATTCCTCTTCCAAAGGGGGCTTACAAACCGCATATTT	66
Db	601	CACAACATCTGTTTGTGAAACAGAAATTCCTCTTCCAAAGGGGGCTTACAAACCGCATATTT	66
Qy	661	ACTGCTTCACAGTCTCAAAAAGAGATCCACAGGAGGCACTGTGTCAGCTGTACAA	72
Db	661	ACTGCTTCACAGTCTCAAAAAGAGATCCACAGGAGGCACTGTGTCAGCTGTACAA	72
Qy	721	GTGGAAGTAAAGATTTTCTTGGGACGTGATGTCTCTCATGATAGACGAGAAAG	78
Db	721	GTGGAAGTAAAGATTTTCTTGGGACGTGATGTCTCTCATGATAGACGAGAAAG	78
Qy	781	AGTGTCTCTTGTGATGTGCTGTAATTAACAATGCACTGTACCTGTGACAGATATATGCA	84
Db	781	AGTGTCTCTTGTGATGTGCTGTAATTAACAATGCACTGTGTACCTGTGACAGATATATGCA	84
Qy	841	AGGTGTTGAAAAAGAAAAATGACTGCAACAAGCTTGAAGCAATTAAGCTTCAATGAGAC	90
Db	841	AGGTGTTGAAAAAGAAAAATGACTGCAACAAGCTTGAAGCAATTAAGCTTCAATGAGAC	90
Qy	901	CAGATTTTATAGGAGCTTCCTAGSAACTCAAAATTAATTAAGTTGATGATGAAGCGCATGGA	96
Db	901	CAGATTTTATAGGAGCTTCCTAGSAACTCAAAATTAATTAAGTTGATGATGAAGCGCATGGA	96
Qy	961	AGGTACCCGAATCCTTTTCTATGATCAGAGAGATTAATTCACATGTTGTGTGTGTGA	102
Db	961	AGGTACCCGAATCCTTTTCTATGATCAGAGAGATTAATTCACATGTTGTGTGTGTGA	102
Qy	1021	TGCTGCACTGAGTGTGCGGCTCTCTCTGAGATCAATTTGTCAATCTTGTACTGTATATTT	108
Db	1021	TGCTGCACTGAGTGTGCGGCTCTCTCTGAGATCAATTTGTCAATCTTGTACTGTATATTT	108
Qy	1081	GTGATTCACCAAGATATAGACTGTAGAGGTATATCATCTTTCTTATAGTGTATATGAT	114
Db	1081	GTGATTCACCAAGATATAGACTGTAGAGGTATATCATCTTTCTTATAGTGTATATGAT	114

Db	1081	GTGATTCAACCAAGAAATATAGACTGTAAGTGATGATCATCTTTCTTTCATGTGATTAAGAT	114
Qy	1141	ATTATTCAGATGATATATATCTTTCAGCTAATATTAATAGAAACATAAGCTTTGCAGC	120
Db	1141	ATTATTCAGATGATATATCTTTCAGCTAATATTAATAGAAACATAAGCTTTGCAGC	120
Qy	1201	CTCACCAAGTGCTCTGTAATTTGGAAGTTCTTAATTTGTCGTTGATTTGAAGATT	126
Db	1201	CTCACCAAGTGCTCTGTAATTTGGAAGTTCTTAATTTGTCGTTGATTTGAAGATT	126
Qy	1261	ACTGATTAATAA 1271	
Db	1261	ACTGATTAATAA 1271	
RESULT 2			
AX093580	Sequence 1 from Patent WO0118150.	1271 bp	DNA
LOCUS	AX093580	1271 bp	DNA
DEFINITION	AX093580	1271 bp	DNA
ACCESSION	AX093580	1271 bp	DNA
VERSION	AX093580.1	1271 bp	DNA
KEYWORDS	GI:13510017	1271 bp	DNA
SOURCE	Solanum tuberosum (potato)	1271 bp	DNA
ORGANISM	Solanum tuberosum	1271 bp	DNA
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiales; Solanales; Solanaceae; Solanum.	1271 bp	DNA
AUTHORS	Enhardt, T., Lerchl, J., Stille, Nigol, M., Zrenner, R. and Schroeder, M.	1271 bp	DNA
TITLE	Dihydrochrocin extracted from plants	1271 bp	DNA
JOURNAL	Patent: WO 0118190-A 1 15-MAR-2001; BASF AKTIENGESELLSCHAFT (DE)	1271 bp	DNA
FEATURES	Location/Qualifiers	1271 bp	DNA
source	1..1271	1271 bp	DNA
CDS	/organism="Solanum tuberosum"	1271 bp	DNA
	/mol_type="unassigned DNA"	1271 bp	DNA
	/db_xref="taxon:4113"	1271 bp	DNA
	9..1049	1271 bp	DNA
	/note="unnamed protein product"	1271 bp	DNA
	/codon_start=1	1271 bp	DNA
	/protein_id="CAC35419.1"	1271 bp	DNA
	/db_xref="GI:13510018"	1271 bp	DNA
	/translation="MELSIQPDMDMHLRLDQDVLKAVSHAHFGRATVMPVLPKPP	1271 bp	DNA
	ITTTAAVAAYREALILKSPVDSDFNPMTLYLDTTSPMEIKLARSESQVGVCLYAR	1271 bp	DNA
	GATNSQDGVTLFGKCLPYLOEMVEMHMLVHGTEPNBMLKARSESQVGVCLYAR	1271 bp	DNA
	LVQEPOLKVMHEHTTIDAVKPYESCSTGCVAAATTVPRHVLNRNSLFGQGIQPHY	1271 bp	DNA
	CLPLKRIHREALEVSAVTSGRKFRFGLTSSAPHDRRKSCGCGACIYNAPVALSY	1271 bp	DNA
	AKVKEKNALDKLEAFTSFNGPDYGLPRNNSSKIKLSKTPMKVPEFSFVASCPIIPMF	1271 bp	DNA
	AGEMLDMLPAFL"	1271 bp	DNA
ORIGIN			
Query Match	100.0%; Score 1271; DB 6; Length 1271;		
Best Local Similarity	100.0%; Pred. No. 4.8e-313; Indels 0; Gaps 0;		
Matches 1271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	TTGCAAAATGAGCTTCAATCAACACACCTGATGATTTGCAATCTTCATCTCCGTATG	60
Db	1	TTGCAAAATGAGCTTCAATCAACACACCTGATGATTTGCAATCTTCATCTCCGTATG	60
Qy	61	GTGATGTTCTTAAGGCGTGTCTCCACAGTGCACTACTTTGGAGGCGAATAGTCA	120
Db	61	GTGATGTTCTTAAGGCGTGTCTCCACAGTGCACTACTTTGGAGGCGAATAGTCA	120
Qy	121	TGCCAAATTTGAAGCGCTCTATCACTACACCTGCTGCTGAGCATACCGGAGGCGA	180
Db	121	TGCCAAATTTGAAGCGCTCTATCACTACACCTGCTGCTGAGCATACCGGAGGCGA	180
Qy	181	TATTTGAATTTTAACTGTTGATAGTGAATTTCAACCTCTTATGACACTTTATTTGACAG	240
Db	181	TATTTGAATTTTAACTGTTGATAGTGAATTTCAACCTCTTATGACACTTTATTTGACAG	240
Qy	241	ATTCACCAAGTCTTATGAAATTAATCACTGCAAGAGAGCCAGGCGATTTGGGGTCA	300

Db 241 ATACAACTCTCTATGAAATCAAACTAGCAGAGAGAGCCAGTCGTATTTGGGTGA 300
QY 301 AGTTGTAACCTGTGTGTCACGACAAATTTCTCAAGATGAGTACTGATCTTTTGGGA 360
Db 301 AGTTGTAACCTGTGTGTCACGACAAATTTCTCAAGATGAGTACTGATCTTTTGGGA 360
QY 361 AGTGTTCACGATTTCAAGAAATGTTAGACATAATATCTCTGCTGCTTCATGAG 420
Db 361 AGTGTTCACGATTTCAAGAAATGTTAGACATAATATCTCTGCTGCTTCATGAG 420
QY 421 AGGTCTATCTCTGAGGTGACATGTTGATGAGAAAAGTATTCATTTGAACGGTTC 480
Db 421 AGGTCTATCTCTGAGGTGACATGTTGATGAGAAAAGTATTCATTTGAACGGTTC 480
QY 481 TAAAGCCGTGTGTCAGAAATTTCCAAATTTGAAGGTGTGATGAGCATGTTACCA 540
Db 481 TAAAGCCGTGTGTCAGAAATTTCCAAATTTGAAGGTGTGATGAGCATGTTACCA 540
QY 541 TTGATGCTGTAAAGTTGTTGTAATCTTGCACTGAAAGATTTGTCAGCAACTGT 600
Db 541 TTGATGCTGTAAAGTTGTTGTAATCTTGCACTGAAAGATTTGTCAGCAACTGT 600
QY 601 CACAACATCTGTTTGAACAGAAATCTCTCTTCCAAAGGGGGCTTCAACCGCATAT 660
Db 601 CACAACATCTGTTTGAACAGAAATCTCTCTTCCAAAGGGGGCTTCAACCGCATAT 660
QY 661 ACTGCTTCCAGTCTTCAAAAGAGATTCACAGGAGGCACTTGTCTAGCTGTAA 720
Db 661 ACTGCTTCCAGTCTTCAAAAGAGATTCACAGGAGGCACTTGTCTAGCTGTAA 720
QY 721 GTGAGATTAAGATTTTCTTCTGGAATCTGATGATGCTCTCTCATGATAGCAAG 780
Db 721 GTGAGATTAAGATTTTCTTCTGGAATCTGATGATGCTCTCTCATGATAGCAAG 780
QY 781 AGGTCTTCTGAGTGTGCTGTATTTACAATGCACTGACCTGTACATATATG 840
Db 781 AGGTCTTCTGAGTGTGCTGTATTTACAATGCACTGACCTGTACATATATG 840
QY 841 AGGTGTTGAAAGAAATGCACTGCAAGCTTGAAGCACTTCACTAGCTTCAATG 900
Db 841 AGGTGTTGAAAGAAATGCACTGCAAGCTTGAAGCACTTCACTAGCTTCAATG 900
QY 901 CAGATTTTATGAGCTTCTTCAAGCACTCAAGATTTAAGTGAAGACGCGCAT 960
Db 901 CAGATTTTATGAGCTTCTTCAAGCACTCAAGATTTAAGTGAAGACGCGCAT 960
QY 961 AGGTACCCGAATCTTTTCTTANGCATCAGAGATTTATCCATGTTGCTGTA 1020
Db 961 AGGTACCCGAATCTTTTCTTANGCATCAGAGATTTATCCATGTTGCTGTA 1020
QY 1021 TGCTGCACTGCTGCTGCTCTCTCTGAGATCATTTTCTTCTTCTGATCTTA 1080
Db 1021 TGCTGCACTGCTGCTGCTCTCTCTGAGATCATTTTCTTCTTCTGATCTTA 1080
QY 1081 GTGATTCACCAAGATTAAGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1081 GTGATTCACCAAGATTAAGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 ATTATCAGATGATTAATCTTCTTCACTAATTAATTAAGAAACAATTAAGCTT 1200
Db 1141 ATTATCAGATGATTAATCTTCTTCACTAATTAATTAAGAAACAATTAAGCTT 1200
QY 1201 CTCACAAAGTCTCTGATTTGAAATGTTCTTAAATTTGTTGATTTTGAAG 1260
Db 1201 CTCACAAAGTCTCTGATTTGAAATGTTCTTAAATTTGTTGATTTTGAAG 1260
QY 1261 ACTGATTAATA 1271
Db 1261 ACTGATTAATA 1271

LOCUS AX470409 1743 bp DNA linear PAT 09-AUG-2002
DEFINITION Sequence 2 from Patent WO2052012.
ACCESSION AX470409
VERSION AX470409.1 GI:22205539
KEYWORDS
SOURCE
ORGANISM
Beta vulgaris
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllales; Amaranthaceae; Beta.
REFERENCE
1 Kanhonou, R.A., serrano Salom, R. and ros Palau, R.
AUTHORS Sugar beet genes involved in stress tolerance
TITLE Patent: WO 02052012-A 2 04-JUL-2002;
JOURNAL CropDesign N.V. (BE)
FEATURES
Location/Qualifiers
1..1743
/organism="Beta vulgaris"
/mol_type="unassigned DNA"
/db_xref="taxon:161934"
ORIGIN
Query Match 50.8%; Score 646; DB 6; Length 1743;
Best Local Similarity 76.4%; Pred. No. 1e-153;
Matches 793; Conservative 0; Mismatches 245; Indels 0; Gaps 0;
QY 3 GCAAAATGAGACTTCTCAATCACACAACTGATGATTTGGATCTTCATCCGGATGT 62
Db 193 GTAAAGATGAGACTTCTTCAACGCGCTGAGACTGGCATCTTCACTCCCGATGA 252
QY 63 GATGTTCTTAAGGCAATGTCCTCAAGTGCATCATCTTTGGAGGCAATAGTCATG 122
Db 253 GATCTTCTGCTGCTGCTGCTCTTCAAGTGCATCATCTTTGGAGGCAATAGTCATG 312
QY 123 CCAATTTGAGGCTCTTCACTCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 182
Db 313 CCGATTTAAGGCTCTTCACTCACTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 372
QY 183 TTGAATCTTACCTGTTGATGATTTCAACCTCTTATGACATTTTGAACGAT 242
Db 373 ATGAGATTAATGCTGATGATGATGATGATGATGATGATGATGATGATGATG 432
QY 243 ACAACGATCTAATGAAATCAAACTGACAGAGAGAGCCAGTCTATTTGGGATG 302
Db 433 ACAGCCAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 492
QY 303 TTGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362
Db 493 TTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552
QY 363 TGTTTACAGTCTTCAAGAAATGTTGATGATTTGCTGCTGCTGCTGCTGCTG 422
Db 553 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 612
QY 423 GTTATCTAATCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 482
Db 613 GTTACAGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 672
QY 483 AGACGTTGTCAGAAATTTCCAAATTTGAAGTGTGATGAGCATGTTTACACCAT 542
Db 673 AGACCTTTAATTCAGAAATTTCCAAATTTGAAGTGTGATGAGCATGTTTAC 732
QY 543 GATCTGTAAATTTGTTGAATCTTGAATCTTGAATCTTGAATCTTGAATCTTGA 602
Db 733 GATCTGTAAATTTGTTGAATCTTGAATCTTGAATCTTGAATCTTGAATCTTGA 792
QY 603 CAATCTGTTTGAAGAGGATTTCTCTTCAAGGGGGCTTACAAACGCAATTAAC 662
Db 793 CAGCACTGTTGTTGAATGAAATCTCTCTTCAAGAGGGGTGCAACGCAATTAAC 852
QY 663 TGCTTCAGTCTCAAAAGAGATTCACAGGAGGCACTTGTGTCAGCTGTAAAG 722
Db 853 TGTCTTCAGTCTCAAAAGAGATTCACAGGAGGCACTTGTGTCAGCTGTAAAG 912

QY 723 GGAAGTAAAGATTTTCTGGAGCTGATAGTCTCTCATGATGACGAAGAAAGAG 782
DB 913 GGGAGACAGCAATATTTCTTGGAGCTGATAGTCTCTCATGAAAGCGAGGAAAGAA 972
QY 783 TCTTCTTGATGATGCTGATATTTACATGACCTGTAGCTTGTCTGATATGCGAG 842
DB 973 TCTTCTTGATGATGCTGATATTTACATGACCTGTAGCTTGTCTGATATGCGAA 1032
QY 843 GTGTTGAAAGGAAATGCACTGACAGCTTGAAGCTTCACTAGCTTCAATGACCA 902
DB 1033 GTATTGAAAGGCTGGGCTGACCAAGTTAGAGGCAATTACAGCTTAAATGACCC 1092
QY 903 GATTTTATGGCTCTCTAGGACAACTCAAGATTAGATTGATTAAGCCCTGGAAG 962
DB 1093 GATTCTATGCTCTTCAAGGAATATCTGCAAGATCAATGAAAAAACAATGAAA 1152
QY 963 GTACCCGATCTCTTTTATGATCAGAGATATTTATCCATGTTGTGTGTAATG 1022
DB 1153 GTCTAGAGCGTATACCTTCCATCCGAGAAATTAATCCCTATGTTGTGTAATG 1212
QY 1023 CTGCACTGTTGCCGCT 1040
DB 1213 CTGACTGGAAGCATCT 1230

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

RESULT 4

BVU251897

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

AUTHORS

1029 TGCTGCCGCT 1040
1021 TGAAGCCATCT 1032
RESULT 5
AF0000146
LOCUS
DEFINITION
ACCESSION
AF0000146
VERSION
AF0000146.1
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1362)
Zhou, L., Lacroix, F. and Thornburg, R.
Characterization of the Arabidopsis thaliana cDNA encoding
Dihydroorotase (Accession No. AF000146) (PGR97-115)
Plant Physiol. 114, 1569 (1997)
2 (bases 1 to 1362)
Zhou, L., Lacroix, F. and Thornburg, R.W.
Direct Submission
Submitted (16-APR-1997) Biochemistry and Biophysics, Iowa State
University, Ames, IA 50011, USA
location/Qualifiers
1..1362
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
<1..>1362
/gene="pyr4"
86..1219
/EC_number="3.5.2.3"
/function="catalyzes the intramolecular condensation of
cardamoyl aspartate to form dihydroorotase"
/codon_start=1
/product="dihydroorotase"
/protein_id="AA071134.1"
/db_xref="GI:2121273"
translating="MIRKIVSPYSGFSQKIKRDSSEKVTAVRMELITPDPMH
LHARDGLHVVHSHSNFKRAIVMPKPTSTAAIYKFKIKALPSESSFDP
LMTLYDKTPEIRILARESGVYAVLYAGATNSODVTLDFKCLPVEEMK
OMPLIVHGEVTPSIDVREDEKIFIEVLQPLIOLPOLKVMHEHITMDAVNFS
CKESVGVATVTPHLLNRNLFGGLOPHNYCIPULKREIHRALYKAVTSKKEP
LGTSDAPHERKSSCGCAGIYAPILALSLYKVFDEAGLDELKLEAFSTFNGDFYIG
LPRNSKITLKRKSPKVPYDVENFPFGEIVPMFAGETLQWOLK"
polyA_signal
1325..1330
/gene="pyr4"

Query Match 46.2%; Score 587; DB 8; Length 1362;
Best Local Similarity 72.9%; Pred. No. 1.2e-138;
Matches 755; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

3 GCAAAATGAGCTCTCATCAACACCACTGATGATGCACTTCATCCGATGATG 62
176 GTAAGAAAGAGCTCAATCACTACCTGATGATGCACTTCATCCGATGATG 235
63 GATGTTCTTAAGGAGCTGCTCTCAAGTGCATCACTTTGGAGGCAATAGTCATG 122
236 GATCTTCTTCAATGCTGTTGTTCCCAAGTGCATTAATTTAAGAGCAATGTGATG 295
123 CCAATTTGAAGCTCTCTATCACTACCTGCTGCTGTAAGCATACCGGAGCGCAT 182
236 CCAATTTGAAGCTCTCTATCACTACCTGCTGCTGTAAGCATACCGGAGCGCAT 355
183 TTGAATCTTACCTGATGATGATTTCAACCGCTTATGACACTTATTTGACAGAT 242
356 ATGAAGCTTTCCTGATGATGATGATTTGATTCACCTTATGACACTTATTTGACAG 415

243 ACAACAGTCTATGGAATCAAACTAGCAGAGAGAGCCAGTCTATTTGGGTGAAG 302
416 AAACTCTACCTGAGAGATCAGGCTTGCAGAGGAAAGTGTGTTATGGGTGAAG 475
303 TTGTACCTCTGCTGTCGCAAGCAATTCAGATGAGTGAATCTTTTGGGAGAG 362
476 GTGTACCTCTGCTGTCGCAAGCAATTCAGATGAGTGAATCTTTTGGGAGAG 535
363 TGTTCACAGTCTTACAGAAATGTTGAGCATTAATGCTCTGCTGCTTATGAGAG 422
536 TGTTCACAGTCTTACAGAAATGTTGAGCATTAATGCTCTGCTGCTTATGAGAG 595
423 GTTACATCTCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 482
596 GTTACATCTCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 655
483 AGACCGTGTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 542
656 CAGCTCTATTCAGAGCGCTTCCGAGTGAAGTGAATGATGATGATGATGATGATGAT 715
543 GATGCTGTAAAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 602
716 GATGCTGTAAAGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 775
603 CAACATCTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 662
776 CAACATCTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 835
663 TGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 722
836 TGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 895
723 GAAAGTAAAGATTTTCTTGGAGCTGATGATGATGATGATGATGATGATGATGAT 782
896 GAAAGTAAAGATTTTCTTGGAGCTGATGATGATGATGATGATGATGATGATGAT 955
783 TGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 842
956 TGCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 915
843 GTGTTGAAAGAGAAATGCACTGACAGCTGACAGCTGACAGCTGACAGCTGACAGCT 902
1016 GTCTTTATGAGGAGGAGTGGCTGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1075
903 GATTTTATGAGGAGTGGCTGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 962
1076 GATTTTATGAGGAGTGGCTGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1135
963 GTACCCGATCTTTCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1022
1136 GTTCCAGACGTTTCAACTTCCATGAGAGATGATGATGATGATGATGATGATGAT 1195
1023 CTGAGCTGTTGCCG 1037
1196 CTTCAATGACACCG 1210

RESULT 6
AK122010
LOCUS
DEFINITION
AK122010
ORYZA SATIVA (japonica cultivar-group) cDNA clone:J033111101, full
insert sequence.
AK122010
VERSION
AK122010.1
KEYWORDS
FLI CNA; CAP trapper.
SOURCE
ORYZA SATIVA (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehharitoidae; Oryzaceae; Oryza.
REFERENCE
1
The Rice Full-Length cDNA Consortium, National Institute of

Db	1105	ACAACTACTTTCTTGGTCTGACAGTGTCTCTCTCATGATTAACGGAAGAAAAGATGCTCCG	1164	/gene="slr0294"	1..480	gene
Qy	791	TGGATGTGCTGATATTTACAAATGCACTGTGATGCTGTCTGATATATGCGAAGTGTTTGA	850	/note="ORF_ID:slr0294"		
Db	1165	TGGTTGTGACGAATATATATAGCGCTCCTGTGGCTCTTCTTTATGCGAAGATATTCGA	1224	unknown protein		
Qy	851	AAAGAAAATGCACTGCAACAGCTTGAAGCATTCACAGTTCATTAAGGACCAAGATTTTA	910	/transl_table=1		
Db	1225	ACAGGCGGCGCTCTTATAGCTAGAGATTTTACAGCTTCAATGGCCCAAGATTTTA	1284	/protein_id="BA18517.1"		
Qy	911	TGGGCTTCCTAGGAACAACCTCAAGATTAAGTGAATGACGCAATGGAAGTACCGGA	970	/db_xref="GI:1653605"		
Db	1285	TGGCTTTCAGAGAAACCTTCAAGATTTGTCCTTGAAGAAAGAGCCCTGGAAGTACTGA	1344	/translation="MTKKEALTFPAGILGKWTAKADARAVALNVPNSDLSQTYKHE		
Qy	971	ATCCTTTTCTATGTCATCAGAGATATATATATCCAGTTTGTGTGTAATGCTCGACTG	1030	ADPKSFHEQHSALMELLIKVSTAKRFQKEPWEATWIDAEAKITDNSQTYA"		
Db	1345	TACATCTCATATATAGTTTCAAGAGATCGTGCTATGTTACTGCAACACCTCGAATG	1404	428..823		
Qy	1031	GTTGCCGCGCTCTTCTGAGAA	1052	/gene="slr0392"		
Db	1405	GCTTCATCCGATCACTTGA	1426	428..823		
RESULT 7						
D90915						
LOCUS	D90915	130001 bp	DNA	linear	BCT 05-JUL-2001	gene
DEFINITION	Synechocystis sp. PCC 6803 DNA, complete genome, section:17/27,					
ACCESSION	D90915	AB001339	BA000022			CDS
VERSION	D90915.1	GI:1653604				
KEYWORDS	Synechocystis sp. PCC 6803					
SOURCE	Synechocystis sp. PCC 6803					
ORGANISM	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.					
REFERENCE	1					
AUTHORS	Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N.,					
TITLE	Sugiyama, M. and Tabata, S.					gene
JOURNAL	Sequence analysis of the genome of the unicellular cyanobacterium					
MEDLINE	Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb					
PUBMED	region from map positions 64% to 92% of the genome					CDS
REFERENCE	DNA Res. 2 (4), 153-166 (1995)					
AUTHORS	2					
	8590279					
	Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y.,					
	Miyajima, N., Hikosawa, M., Sugiyama, M., Sasanoto, S., Kimura, T.,					
	Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Nanno, K.,					
	Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A.,					
	Yamada, M., Yasuda, M. and Tabata, S.					
	Sequence analysis of the genome of the unicellular cyanobacterium					
	Synechocystis sp. strain PCC6803. II. Sequence determination of the					
	entire genome and assignment of potential protein-coding regions					gene
	DNA Res. 3 (3), 109-136 (1996)					
JOURNAL	97061201					
MEDLINE	3 (bases 1 to 130001)					CDS
PUBMED	8905231					
REFERENCE	Tabata, S.					
AUTHORS	Direct Submission					
TITLE	Submitted (28-JUN-1996) Satoshi Tabata, Kazusa DNA Research					
JOURNAL	Institute, The First Laboratory for Plant Gene Research, Yana					
	1532-3, Kisarazu, Chiba 292-0812, Japan					
	(E-mail: tabata@kazusa.or.jp, URL: http://www.kazusa.or.jp/cyano/					
	Tel: 81-438-52-3933 (ex.2330), Fax: 81-438-52-3934)					
	Potential protein coding regions were assigned on the basis of					
	similarity search of the ORFs and Genemark analysis.					
FEATURES	Location/Qualifiers					
source	1..130001					gene
	/organism="Synechocystis sp. PCC 6803"					
	/mol_type="genomic DNA"					
	/strain="PCC6803"					
	/db_xref="taxon:1148"					CDS
	/note="by:nygm:Synechocystis PCC6803"					
gene	1..480					
	/gene="slr0294"					
	/note="ORF_ID:slr0294"					
	unknown protein					
	/codon_start=1					
	/transl_table=11					
	/protein_id="BA18517.1"					
	/db_xref="GI:1653605"					
	/translation="MTKKEALTFPAGILGKWTAKADARAVALNVPNSDLSQTYKHE					
	VIEDVTDIDIVLSIVLADPFGVILIDRQKQAAQGSVTKKKIKIEIBSHQEI					
	ADPKSFHEQHSALMELLIKVSTAKRFQKEPWEATWIDAEAKITDNSQTYA"					
	428..823					
	/gene="slr0392"					
	428..823					
	/gene="slr0392"					
	428..823					
	/gene="slr0392"					
	428..823					
	/note="ORF_ID:slr0392"				</	

[illegible]

```

/codon_start=1
/translation="MLFYILCRLLKPTSPRNOASFNQKNNMPNASTLTKKALINM
/db_xref="GI:1653614"
/db_xref="GI:1653614"
/translacion="MLFYILCRLLKPTSPRNOASFNQKNNMPNASTLTKKALINM
KELSLPRRETRAKACGYSTSKESQGVAVNLDPIDVAIAKQVLPDSGRDKDGRRE

Query Match
Best Local Similarity 27.2%; Score 345.6; DB: 1; Length 130001;
Matches 608; Conservative 0; Mismatches 389; Indels 9; Gaps 2

QY 13 AGCTCTCATACACACACTGATGATATTGGCACTTTCATCTCCGATGATGATGTTCTTA 72
Db 20191 AGCTTACCATCACTCGACCCACACACATCGGACCTCGCATCTACGGGATGAGCGGCCCTGA 202
QY 73 AGGCAGTTGTCCTCACAGTSCGACATCACTTGGGAGGGGGAATAGTCATGCCCCAATTTGA 132
Db 20251 AAGCAGTTCTACCCACACTGTCTCCGTAAATTTGCTAGGGGAGATGTCATGCCCAATTTGA 203
QY 133 AGCCCTCCATATCACTACCACTGCTGCTGTGTAGCATACCGGAGGCGATATGGAATCTT 192
Db 20311 AGCCTCCGTCGCTTCCGTGGCTGATGCCGCCCTATCGGGAAAGATTCGCCGCCA 203
QY 193 TACCTGTGTAGTATGATTTCAACCCCTCTTATGACATTTATTTGACAGATACCAAGTC 252
Db 20371 TACCTGGGGAGGCGCAGTTTGAACCCGCTATGACCTTTATCTAACGGATATATCTAAATC 204
QY 253 CTATAGGAATTCACAACTAGCAAGAGAGACGCCAGTGTATTTGGGGTGAAGTTTACCTCG 312
Db 20431 CCGAAGAAATTTATGGCGCCAAAGCAATCCAGTTGTATTAACGATTAATACTATCCCG 204
QY 313 CTGGTGCACAGCAAAATCTCAAGATGAGATGATGATCTTTGGGGAAAGTTTACAG 372
Db 20491 CTGAGGCAACCAACCATCTGACTTTGGGTACCGA---TATCATGATGTGATGACG 205
QY 373 TTCTACAAAGAAATGSGTTGAGCATATATGCTCTGCTGTCGTATGAGAGAGTTACTAATC 432
Db 20548 TGTGGCAGCCATGGAACAGGTGACCTTGCCTTATCTCACGCGCAGAAATTAACGACA 206
QY 433 CTGAGGTGACATGTTTGTATGAGAAAGATTCATTTGAAACGGTTTAAACCGTTG 492
Db 20608 GTATATTTGATATTTTTCGACGAGAAAGATTCATTTGAAATATTTAATTTCTCTCA 2066
QY 493 TGCAGAAATTTTCACAAATTTGAAGTCGTGATGAGACATGTTACCAACATGATGCTGTA 552
Db 20668 GGGAAAAATTTCCCGCGCTGGGGTAGTGTGGAACATATTACTAATTCAGATCAGATGTC 2072
QY 553 AGTTTGTGATCTTTCGACTGAGAGATTTGTTGACAGAACTGTACCCCAACATGTTG 612
Db 20728 AATTTGTTTTTATCTGCTAAATA-----TATTCAGCGACGATTAAGCCCCCAATTTAC 2078
QY 613 TTTTGAAAGGAATTTCTCTTTCACAAAGGGGCTTACAAACGCAATATTTACTGCTTCCAG 672
Db 20782 TATTTAATGCGAACGCTTTATTTCAAAGGTGCAATTTGTCTCCCACTTTATTTGCTCCAA 2084
QY 673 TCTTCAAAAGAGATTCACAGGAGGCACTTGTGTCAAGCTGTATGACAGTGAAGTAA 732
Db 20842 TTTGAACCGGGAGAAACATCGTTTATCATTTGCTAACCGGCGCAATCTGTGATTCCTA 2090
QY 733 GATTTTTTCTTGGACATGATAGTCTCTCTATATGAGAAAGAAAGAGTGTCTTGG 792
Db 20902 AATTTTTTCTCGGACATGATAGTGTCTCTCCACGCTGCAATATGTAAGAGAGTTATGTG 2096
QY 793 GATCTGCTGATTTTACATGCACTGTAGCTTGTCAATATATGCAAGTGTGTA 852
Db 20962 GCTGTGCTGTGTCTATTTCTGCTCTCCATCATGATGAGAGTATATGCGAAGCTTTGAAA 2102
QY 853 AGGAAAAATGCACTGCAGAGCTTGAAGATTCATGATCTTCAATGACAGATTTTATG 912
Db 21022 GTGTATATTCCTTGATTAAGCTAAGAGCTTTGCTAGCTTTATGAGCCAGATTTTATAC 2108
QY 913 GGCTTCTTAGAACACTAAAGATTAAGTTAGTAAAGCCATGGAAGATCCGCAAT 972

```


/protein_id="CAD84482.1"
/db_xref="GI:30180180"
/translation="MERKRVGIFVALVTLTSLINQPATAGRTVEDPQVNGNITLALGNFG
FVNBGNDLKFRWMEGQRFENDSSQFTQAIIRGLQYATITDKIIWAGYAMIPSD
ELVFPVSGLEFDESHRIWOQVYADNDEFSGKLSIRSFEDRPFDDHNAFVSGDDVAYR
KQIVKKAIPVAMIDPNLFTIIONELFTIGNTVSNPQFISRGFDQNRAPFGLGYKHON
ATVELGMNQFIDRRNRPDPDQNMHNFAVNLFLN"
misc_feature
/locus_tag="NE0571"
/note="Signal predicted by SignalP 2.0 HMM (Signal peptide
probability 0.999) with cleavage site probability 0.983 at
residue 25"
/gene="CD1"
/locus_tag="NE0572"
complement(6048..6965)
/gene="CD1"
/locus_tag="NE0572"
/function="InterPro IPR000847:IPR005119 COG6 COG0583"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Bacterial regulatory protein, LysR family"
/protein_id="CAD84483.1"
/db_xref="GI:30180181"
/translation="WNFOQLKISSTIRONYLTENVALLTSQSGVSKIKDLEDEL
NIDLFTRKGRITGLTAPGKELVKTIVRIIDTNNIRLAEQFSOHQOGLVTATHT
QAVYALPDVVRQFKEIPKVALVHSESPGRIIVMLNGADIGIATEALVELTS
PEYVYTMHAIIVPAHPLQSIHPLTLEVAQFPIITVHKPFGSRIDQAFSAELRP
NIAMGALDADYIKVTYVLMGIGIVASVAFPRPDYTKIKIDSHLEFQNTCISIR
NHLYRGVYRFKEICIPSLSRDITCSERQ"
complement(7060..8685)
/locus_tag="NE0573"
complement(7060..8685)
/locus_tag="NE0573"
/function="InterPro IPR000192 COG6 COG0520"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="aminotransferase class-V"
/protein_id="CAD84484.1"
/db_xref="GI:30180182"
/translation="MNTNEFNQWIDKVGSAVYVDISILNMGNALFNTPIRST
EPPALVPGAEVSHSRPATGSSYPTIPLCEALFLTPNOYRAPAISDSHS
GHPWESSFVFLDEGIAPEHRRDAOIPIAHNRPDIHARDFPIQEHVNGSLVWL
DNATITQPOAVIDRLTYETTHNSNHRRAHELAATDAEESNKROGLNAST
DEILFAGTEBAINLVAAQSGKROHISAGDEIITVLEHNANIVPMQLCNETGARLRV
VPURDQVULLEIYOKLNSHRLVAFSQVSNALGITPARQVEMAHNVEGARVLDG
AQSVSHRVVQQLDCDWVFVSGHKVFGPGIGIVLIGKTELNDOPMGCGGMMIODY
TFEKTAHYGAPAREAGTGNIDAVGLGAIIDYVERIGLENTSRVHEIMTYATGCKL
SIGRLRIGTAPKAGVLSFTLKDPSFTEVGTALNREGIAYVAGHNCAPILIRFGVE
STVRPSLAFFNTYTDIDRLAAATDIRKKNAP"
gene
complement(8672..9589)
/locus_tag="NE0574"
complement(8672..9589)
/locus_tag="NE0574"
/codon_start=1
/codon_tag="NE0574"
/evidence=not experimental
/transl_table=11
/product="major membrane protein I"
/protein_id="CAD84485.1"
/db_xref="GI:30180183"
/translation="MTDHIQDTALGDVAARTLANATKTIIPWGGITPRMLTHLLQWV

Query Match 23.1%; Score 294.2; DB 1; Length 306050;
Best Local Similarity 57.4%; Pred. No. 4e-64;
Matches 593; Conservative 0; Mismatches 428; Indels 12; Gaps 3;

QY 13 AGGCTCAATCAACAACCTGATGATGGCATCTTCCTCGTGATGATGATGATCTTGA 72
DB 171791 AGCTGACCTTTTACCGCTCGGATGACTGCTGATTCCTGATGACGATGACGATGAC 171850
QY 73 AGGCAATGCTCTCAACATGATCACTTTGGAGGCAATAGTCAATGCAAAATTTGA 132

DB 171851 GATCGAGTGTGCCGATACCGCCGCGCTTTTGCGGCGCCATCATGCGCAACTCA 171910
QY 133 AGGCTCTTATCACTACCACTGCTGCTGCTGATGACATACCGGAGGCGATATGAATCTT 192
DB 171911 AGCGGCCCATCGTACGACTGAAACAGGCGAGTCCGCTATGTCGTGTCGTT 171970
QY 193 TACCTGTGA-----TACTGATTTCAACCCCTTATGACACTTTATTTGACAGATCA 246
DB 171971 TACCGGCTAGCTGCGCGCGCGCGCTTGACCGCTATGACGCTATGACCTGACGATACGA 172030
QY 247 CCAATCTTATGAAATCAAACTAGCAAGAGAGAGAGAGAGAGATTTGGGGTGAAGTGT 306
DB 172031 CATCACCCGAAATAATCTCGGGCTCAGGCAAGTGTATGTGCAAGTATCAAGCTCT 172090
QY 307 ACCCTGCTGATGCAACGAAATTTGCAAGATGAGTACGATCTTTTCGGAAGTGT 366
DB 172091 ACCCTGCTGAGCGACACACATTCGATGCGGGGTGTCACGATATTT---GCACGCTGCG 172147
QY 367 TACCAATCTTCAAGAAAATGCTTGAAGCATTAATGCTCTGCTGATTCAGAGAGGTTA 426
DB 172148 AGGCTACGCTGAGAAATGGAAGGCTGACATGCTTTATCTGCTGATGAGAAATGG 172207
QY 427 CTATCTGAGGTTGACATGTTTATAGAGAAAAGTATTCATTTGAACGGTCTTAAGAC 486
DB 172208 TCGATCTGCAAGTGAATGTTTTCACCGGAAAAATTTTCATTGACCGGGTACTGACGC 172267
QY 487 CGTTGGTCAAGAAATTTCCACAAATTTGAAGGCTGATGAGACATGTTACCAACCATTTGATG 546
DB 172268 CGCTACTGCAACGCTTCCGGGCTGGGAGTGGATTCGAAATATCATCATCAACAGAGAG 172327
QY 547 CGTTAAATTTTGTGAATCTTGACACTGAAGATTTGTGACGACACTGTGACCCCAAC 606
DB 172328 CAGTCAAGTGTGTCAAG---GGCTCAAAACCGGATTCGTCGACGATTAACGACAC 172384
QY 607 ATCTGTTTGAACAGAAATCTCTCTTCAAGGGGCTTACAAACCGCATTAATTTACTGCG 666
DB 172385 ACTGATGCTGAACCGGAATGCCCTGTTACCGGAGATTCGCTCGCATCATTTACTGTT 172444
QY 667 TTCCAGTCTCAAAAGAGATTCACAGGAGGCACTTGTCAGCTGTACAAAGTGA 726
DB 172445 TCCGAGTACGAAACGGGAAATACATGTCAGGCACTGCTGAAAGCAGGACCAAGCGGTC 172504
QY 727 GTAAGATTTTCTTGGAGCATGATAGTCTCTCATATATAGACGAAAGAAAGTGT 786
DB 172505 ATTCGAATTTCTTCTTGATACGACGACGACACATCCGCTCAGAGCAAGATCAG 172564
QY 787 CTGTGATGTGCTGATTTTACAAATGCACTGTAGCTTGTCAATATGGAAGTGT 846
DB 172565 CCGTGGGTGTGCGGGTATTTACAGTGTCAATGAGCCATGAAATTTATGCGGAGTAT 172624
QY 847 TTAAGAAAGAAATGCACTGCAAGCTGGAAGCATCTACTAGCTTCAATGACCAAT 906
DB 172625 TCGAAGAGCTGCGCGCTGATCGACTGAGAGCTTTTACCAAGCTTTCACGCGCCCGAT 172684
QY 907 TTATGAGCTCTCTTGAAGCAACTCAAGATTAAGTGAAGTGAAGGATGAC 966
DB 172685 TCTATGTTTACACAGATATACCAATCGAATTTCCCTATGACGGAAGCTGCGAGATTC 172744
QY 967 CCGAATCTTTTCTTATGATAGAGATTTATTTCCATGATTTGCTGTGTAATGCTCG 1026
DB 172745 CGGAGAGGTGCAATTTGGGTGAAGAAATATGATCCCTCAGAGAGGAGGAAACGCTG 172804
QY 1027 ACTGTTGCCGCG 1039
DB 172805 GCTGGAATGCG 172817

RESULT 9
AP005085/c 303450 bp DNA linear BCT 07-Apr-2004
LOCUS AP005085
DEFINITION Vibrio parahaemolyticus DNA, chromosome 2, complete sequence, 2/6.
ACCESSION AP005085 BA000032

VERSION	AP005085.1	GI:28808465
KEYWORDS	Vibrio parahaemolyticus	
SOURCE	Vibrio parahaemolyticus	
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.	
REFERENCE	1	
AUTHORS	Nasu, H., Iida, T., Sugahara, T., Yamaichi, Y., Park, K. S., Yokoyama, K., Makino, K., Shinagawa, H. and Honda, T.	
TITLE	A filamentous phage associated with recent pandemic Vibrio parahaemolyticus O3:K6 strains	
JOURNAL	J. Clin. Microbiol. 38 (6), 2156-2161 (2000)	
MEDLINE	20295086	
PUBMED	10834969	
REFERENCE	2	
AUTHORS	Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K., Uda, T., Tagomori, K., Iijima, Y., Najima, M., Nakano, M., Yamashita, A., Kubota, Y., Kimura, S., Yasunaga, T., Honda, T., Shinagawa, H., Hattori, M. and Iida, T.	
TITLE	Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V. cholerae	
JOURNAL	Lancet 361 (9359), 743-749 (2003)	
MEDLINE	22508454	
PUBMED	12620739	
REFERENCE	3	
AUTHORS	(bases 1 to 303450)	
TITLE	Hattori, M., Yamashita, A., Oshima, K. and Shiba, T.	
JOURNAL	Submitted (09-APR-2002) Masahira Hattori, Kitasato Institute for Life Sciences, Kitasato University, Kitasato 1-15-1, Sagamihara, Kanagawa 228-8555, Japan	
MEDLINE	(E-mail:hattori@genome.1b.kitasato-u.ac.jp/, Tel:81-42-778-8194, URL:http://genome.1b.kitasato-u.ac.jp/, Fax:81-42-778-8193)	
PUBMED	16-may-2002	
REFERENCE	4	
AUTHORS	Katsushi Yokoyama, Takayuki Uda, Kenichi Tagomori, Masatomo Najima, Masayuki Nakano, Yoshino Kubota, Shigenobu Kimura, Teruo Yasunaga, Takehi Honda, Hideo Shinagawa, Tetsuya Iida (Osaka University), Yoshio Iijima (Kobe Institute of Health), and supported by the Research for the Future Program of the Japan Society for the Promotion of Science. This clone was isolated from a patient presenting with acute gastroenteritis.	
COMMENT	Location/Qualifiers	
FEATURES	1. .303450	
SOURCE	/organism="Vibrio parahaemolyticus"	
	/mol_type="genomic DNA"	
	/strain="O3:K6"	
	/sub_strain="RIMD 2210633"	
	/db_xref="taxon:670"	
	/chromosome="2"	
gene	81. .1574	
	/gene="VPA0300"	
CDS	81. .1574	
	/gene="VPA0300"	
	/note="similar to EMB:CA646187.1 (A1591787) percent identity 68 in 487 aa"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="hypothetical signal peptide protein"	
	/protein_id="BAC61643.1"	
	/db_xref="GI:28808466"	
	/translation="MNLKAVPKALVLTLLASLPISVAVYAAPNANSSLVQOQMEGNWPTONEAKLVQELVYQRAISAYQOMQANNAVIGRDSSEKFGYVLPINRDM SRTLVPPNADVVVSMNLTLDKETGPLVAAAPVIGMTDFPQRTITDVLIGPDA RGGVLLIPNVDPVPKGYPTFTSTNNVLPFTYWKKGNGADSDPSDKIAEOR IYPLADEKNIQPMKFTPNASGQRINMAYPTPTFSWKKQGVDEEPIASISPERGVU ASIGIVKGFNPNTVEEKQLKEKVKTPKMTIQRLGKDEKVLTKDQYRANA GATSBYQESYLDIDQRAFPQYVSSAPAAVMTNAGSKYPTVSDSKDILLGSH QYKLIHPANPPAKLFWATLVNVVDTGMPETSOILPSSKMGPDNIETKTDGSDIDYFGP QLPNGAPESNYIKTIPGRDPLAIRLVGTGIEGFPDQTKPDDVLKIK"	
gene	complement (1835. .2314)	
	/gene="VPA0301"	
CDS	complement (1835. .2314)	

gene	/gene="VPA0301"	
	/note="similar to DBJ:BAB43758.1 (AP003138) percent identity 57 in 158 aa"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="putative peptide methionine sulfoxide reductase"	
	/protein_id="BAC61644.1"	
	/db_xref="GI:28809467"	
	/translation="MMKITYPAGGCLMGVQPMRHLPGVISTEAGRANKTNDTQSEY DQYACVARTFDPQSIVETLMGFPEIIPYSINKGDDVEKRTGYSPDTKALE IAKSFATRYDQDKIAVEVLPKNVYPSDEHQRLLTRPNDYCHPLDLKHYKWK"	
gene	2493. .2882	
	/gene="VPA0302"	
CDS	2493. .2882	
	/gene="VPA0302"	
	/note="similar to GB:AA64344.1 (L16865) percent identity 52 in 129 aa"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="BAC61645.1"	
	/db_xref="GI:28808468"	
	/translation="MKIEHVALWTERLELKGFEYKFNVAVNSDKYHNPKQFPSSYFL SRSRGARLEIMGECVTCNCNSHQVTLAHPAFALGSEQAIVDQITKTLVEDGYQRI DGPRTYGDGYTSCVLDPPGNRIELTV"	
gene	2974. .3714	
	/gene="VPA0303"	
CDS	2974. .3714	
	/gene="VPA0303"	
	/note="similar to GB:AAL20622.1 (AB008775) percent identity 54 in 243 aa"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="putative regulatory protein"	
	/protein_id="BAC61646.1"	
	/db_xref="GI:28808469"	
	/translation="MNPQNEIILQVNDKRKVOTELSDIGVSGVTIRQDNLFLBQ GYLKRGAGATALQSDIDITRLRVRDIDQTLNKAADVAPEVTVIISGSANALIA RTLAERGDVITITPSAYIAHLIRNTSANIILGQVVOHQEESLVGPTKLCINIHFS TAPLIGIDGPHQDGTFSRDMRADIAEIALAKKRRNIVLTDSSKFGQIYSSIKTIVE ISVLTIDKAPKPSDLQLEKLGVEVVLG"	
gene	complement (3733. .4089)	
	/gene="VPA0304"	
CDS	complement (3733. .4089)	
	/gene="VPA0304"	
	/note="similar to GB:AAF96321.1 (AB004376) percent identity 76 in 117 aa"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="conserved hypothetical protein"	
	/protein_id="BAC61647.1"	
	/db_xref="GI:28808470"	
	/translation="MHQHEKLVNVEGTNIGATKAPFEHFVFGMQFVDYDPAABSG QGIDGGRFYSABEQVSQTSNGAALLVFYSSDIHATLEKAKRGQIIRIFEPGGRPH FTSPSGNEPAAVSESH"	
gene	4319. .5860	
	/gene="VPA0305"	
CDS	4319. .5860	
	/gene="VPA0305"	
	/note="similar to REF:NP_231225.1 (NC_002505) percent identity 83 in 510 aa"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="catalase"	
	/protein_id="BAC61648.1"	
	/db_xref="GI:28808471"	
	/translation="MVSTHWRATQMSKSFLLITVGLASTSLQAGTLTRDNGAPVGN QNSTAGENGSVLLQDVHLIOKQRFARERIRPERVVARAGTGAHGEFVAGDFSLTV"	

gene	5962. . 6429	/gene="VPA0306"	5962. . 6429	/note="similar to GB:AAF42465.1 (AE004235) percent identity 63 in 152 aa"	percent
CDS		/codon_start=1	/evidence=not_experimental	/transl_table=11	
		/product="AnkR protein"	/protein_id="BAC61649.1"	/db_xref="GI:28808472"	
		/translation="MKNVPLLSVLFSLGSPSLFASDRLKETGYGVLELPDARI			
		GNNEVTFVSGFPIDIDNMSSTALMVAAYQGNKDIVRLILDSGNACLDKRNRT			
		ALMGALIKREISIALDLVAECSPDLRNKAGLDLKTALAIYQGEVLKSHQNK"			
		complement(6548. . 7201)			
gene		/gene="VPA0307"	complement(6548. . 7201)		
CDS		/note="similar to GB:AAF42465.1 (AE002563) percent identity 47 in 200 aa"			
		/codon_start=1	/evidence=not_experimental	/transl_table=11	
		/product="putative pyrazinamidase/nicotinamidase"			
		/protein_id="BAC61650.1"			
		/db_xref="GI:28808473"			
		/translation="MTVRVNYQTTAVIDVDPKGFSEICPDLEPVAGALDIYBELKN			
		HAKGKLKIVSRDLHPKAAWAEETPAMLEPKGLPNVDYKRNHRCVLGTTGVELLDGL			
		PVLVLDYPOVKNQMDPDHHPKGYGFHFDVADTKTGAEEFLKCNKIDITVVGGLADLFC			
		VKSSWQALDLCGRKYIVLALTRAVLPDVTDSVIAEMKRGALFVNADDIIVERFA"			
		complement(7234. . 7959)			
gene		/gene="VPA0308"	complement(7234. . 7959)		
CDS		/note="similar to GB:AAF96011.1 (AE004352) percent identity 60 in 241 aa"			
		/codon_start=1	/evidence=not_experimental	/transl_table=11	
		/product="conserved hypothetical protein"			
		/protein_id="BAC61651.1"			
		/db_xref="GI:28808474"			
Query Match	22.9%	Score 290.6;	DB 1;	Length 303450;	
Best Local Similarity	57.4%	Pred. No. 3.3e-63;			
Matches	584;	Conservative 0;	Mismatches 424;	Indels 9;	Gaps 3;
QY	15	CTTCGAACTACACAACTGATGATGGACATCTTCATCTCCGAGAGGATGCTTTAAG	74		
Db	104344	CTGACTATTACTCGTCTCGACACTGGACGTTCACTACGGAGAGGAAAGTTCTAAAA	104285		
QY	75	GCAAGTGTCTCTACAGTGCATATCATCTTGGAGGGCAATAGTATCCAAATTTGAAG	134		
Db	104284	GACACCGTTGGCGATATCAAGCCGCTTCAATATGCCGCGCGCTGATATCCAAACCGTC	104225		
QY	135	CTTCTATCACTACCACTGCTGTGTGTAGGATACCGGAGGGCGGATTTGAAATCTTTA	194		
Db	104224	CCAAGTGTCTCAATACCGAAATGGACATCTGCTTCCCGGATTCGAAATTTGAAAGACAG	104165		
QY	195	CCGTTGATAGTATTTCAACCTCTTATAGACATTTATTTGACAGATACACCAAGCTCT	254		
Db	104164	CACGGGCA---ACAATTCGAACCTCTGATGGCGCTTAACTTAAACGATTAACACACGCCA	104108		
QY	255	ATGAAATCAAACTAGACAGAGCCAGGTCGATTTGGGGGTAAGTTGTATCCCTGCT	314		

[illegible]

AUTHORS	Vasconcelos,A.T.R., de Almeida,D.F., Almeida,F.C., de Almeida,L.G.P., de Almeida,R., Goncalves,J.A.A., Andrade,F.M., Antonio,R.V., Aratipe,J., de Araujo,M.F.F., Filho,S.A., Azevedo,V., Batista,A.J., Batista,L.A.M., Batista,V.S., Belo,A., Vander Berg,C., Blamey,J., Bogo,M., Bonato,S., Bordignon,J., Brito,C.A., Brocchi,M., Butley,H.A., Camargo,A.A., Cardoso,D.D.P., Carneiro,N.P., Carraro,D.M., Carvalho,C.M.B., Cascardo,J.C.M., Cavada,B.S., Chaveir,L.M.O., Pasa,T.B.C., Duran,N., Fagundes,N., Falcão,C.L., Fantinatti,F., Farias,I.P., Felipe,M.S.S., Ferrazi,L.P., Ferro,J.A., Ferro,M.I.T., Franco,G.R., Freitas,N.S.A., Furian,L.R., Gazzinelli,R.T., Gomes,E.A., Goncalves,P.R., Grangeiro,T.B., Grattapaglia,D., Grisard,E.C., Guimaraes,C.T., Hanna,E.S., Hungria,M., Jardim,S.N., Laurino,J., Leoi,L.C.T., Passarella,L., Lima,A., Loureiro,M.F., Lyra,M.C.P., Macedo,M., Madeira,H.M.F., Manlio,G.P., Maranhao,A.Q., Martins,W.S., de Mauro,S.M.Z., de Medeiros,S.R.B., Meisner,R.V., Menck,C.F.M., Moreira,M.A.M., Nascimento,F.F., Nicolas,M.F., Oliveira,J.G., Oliveira,S.C., Paixao,R.F.C., Parente,J.A., Pedreira,F.O., Pena,S.J.D., Pereira,J.O., Perreira,M., Pinto,L.S.R.C., Pinto,L.S., Porto,J.I.R., Potrich,D.P., Neto,C.E.R., Reis,A.M.M., Rigo,L.U., Rondinelli,E., dos Santos,E.B.P., Santos,F.R., Schneider,M.P.C., Seunarez,H.N., Silva,A.M.R., da Silva,A.L.C., Silva,D.W., Silva,R., Simoes,I.C., Simon,D., Soares,C.M.A., Soares,R.B.A., Souza,E.M., Souza,K.R.L., Souza,R.C., Steffens,M.B.R., Steindel,M., Teixeira,S.R., Umenyi,T., Vettore,A., Wassem,R., Zaha,A. and Simpson,A.J.G.
TITLE	Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ 25651070, Brazil
JOURNAL	
FEATURES	Location/Qualifiers
source	1. 208524 /organism="Chromobacterium violaceum ATCC 12472" /mol_type="genomic DNA" /strain="ATCC 12472" /db_xref="taxon:243365" complement(96..563) /gene="p1B2" /locus_tag="CV4209" complement(96..563) /gene="p1B2" /locus_tag="CV4209" /note="identified by sequence similarity; putative; ORF located using GeneMark/Blatx/COG2165" /codon_start=1 /transl_table=1 /product="type IV pilin" /protein_id="AA061869.1" /db_xref="GI:34105515"
gene	/translation="MKQRKQGFLLIELMIVAIGITLAIAPAYODYTKARVSEGLALADAKTAVTEYATNNAMAGTAPFNTTGLAASITGNSVSNIVLANGVQQLAYNSTVNGALVDIVPTVSAGSVQVYCTYGTGTAIASQLSANWVSTCRQ"
CDS	838..2184 /gene="b10A" /locus_tag="CV4210" 838..2184 /gene="b10A" /locus_tag="CV4210" /EC_number="2.6.1.62" /note="identified by sequence similarity; putative; ORF located using GeneMark/Blatx/COG0161" /codon_start=1 /transl_table=1 /product="adenosylmethionine-8-amino-7-oxononanoate transaminase" /protein_id="AA061870.1" /db_xref="GI:34105516"
gene	/translation="MSNQAWLERSFAAVHPTCTQMKRHERLPIVPIASADGWLITDPDGKRYLDGVSSWVNLFGHGPRIKQAVKQKQDSEHWLAGFTRPVVELSERLAALSGLGHAFYSGDASATEIALKMSFYMKVGRPOKTRFVSLNSYHGELVAGLAVTDPVLFSSTYATLKRGLRAPSDDARQAVPGERADVALRAARSLNLLAKGGEIAIVIVSPLVQAGAKMAHDPYILTEIRLDCQYVHLIADENIVGRTGSPFAYQOAGITPDP LCLSKGITGFLPLSCVLTIRDEIITQAFTHDVARGFLHSHSYTGNALACAAALAVLDI
CDS	2347..3816 /locus_tag="CV4211" 2347..3816 /locus_tag="CV4211" /note="identified by sequence similarity; putative; ORF located using Glimmer/Blatx"
gene	/codon_start=1 /transl_table=1 /product="conserved hypothetical protein" /protein_id="AA061871.1" /db_xref="GI:34105517"
CDS	/translation="MRQARPGVLVAPGPRTPDGRALQAAQLSEAGPILSAGERS PARRGRILAONNAAGCQOAVRQVPGIQAAPRILPPAAGARTTAQGAARVDDGGAALPEHPAPKLSASRLLAGLPPVAYAGAAARLACGAVARRRRAGHLLPASAARSRHDPVAVRRPAIGATLRCGATGDAVAGCPDGPGLCGPERGCPALISACQIAVPARLLPARGTRAGRDETPGTTAODRGGRRLSGARRSPAAEPGGMVAAPTOACPGIGRDGGDHRRRGRLAPRQRQLAARRRRRREORPPAPARLLPDCROPERIMLPARPSAALARRGAVAGPARPAGRVSVYSAIGIDALRQSGGMRRDAGPAAAAGGAGHHPQRQPFPTGLAARORPAGAAVAPRSLQPVARVQAGRRKRAATGPRQAAASALPPFPADGIPQORALITMKRRASFVTLPSL"
gene	3845..4963 /gene="prob" /locus_tag="CV4212" 3845..4963 /gene="prob"
CDS	/locus_tag="CV4212" /EC_number="2.7.2.11" /note="identified by sequence similarity; putative; ORF located using Glimmer/Genemark/Blatx/COG0263" /codon_start=1 /transl_table=1 /product="glutamate 5-kinase" /protein_id="AA061872.1" /db_xref="GI:34105518"
gene	/translation="MHSVTHASDRIVVKKVSSLVNTDGRGLDINALARWAEVIELRRKQVYIVSSGAIABGCRLGWTRPKGHEIQLAAAAGVGMGCOAYESAFFRGART AQTLLHEDLADRTYRLNARSTITSLNANVPIINENDVATSEIRFGNDTLGALVTNLIEDALVITLDQRLYSADPRKHPDNEFIHQAEGRDLSDMAGGASSVGTGGMITKILIAKRAARSGAATVIASGRPEPLSLAAGEAIGTQLVATNRMARAKWLADHLKLAGRLLDGGAALAIRGTSLPLPGVSAVGDPLRGEAVACVDAAGHEVARGLVNVYSDERARQIMRKSRTREIBALGYIVEPELIHRDNVAL"
CDS	5057..6130 /gene="pofF1" /locus_tag="CV4213" 5057..6130 /gene="pofF1" /locus_tag="CV4213" /note="identified by sequence similarity; putative; ORF located using Glimmer/Genemark/Blatx/COG0687/TC:3.A.1.11.2"
gene	/codon_start=1 /transl_table=1 /product="polyamine ABC transporter system, substrate-binding protein" /protein_id="AA061873.1" /db_xref="GI:34105519"
CDS	/translation="MNKIALAALLAVACGSAAEVDLHIVNNNALSPDTAKPEQSC KCTLVQDYGDNEEMLAKLAAGKGYDMVPTFAVNTLLKQGLQPLDYSRLPNWNL LNSGYALNPPQGNRYAPATVYSLTLIGYNTOLKKAEGVSGNSMALVFPDALLAKIKGKTYVDSQSEIMAAALMYGCKDANSNPPADWKAARETIRAKRYMAFNNQSYIKEILTVGINIWAALGSDNLFQAOQDPAKAKRPFELATYPOEGNVLAIDNNITLKDARPDLAKHINFMGLKQKASEISNOIGATNPVKAABAFKPOIKANPVITMLDPSKQYVALKDLVYSRRELRLMSQVKR"
gene	6173..6757 /locus_tag="CV4214" 6173..6757 /locus_tag="CV4214"
CDS	/complement(6173..6757) /locus_tag="CV4214" /note="identified by sequence similarity; putative; ORF located using Glimmer/Genemark/Blatx/COG3575" /codon_start=1 /transl_table=1

```
/product="conserved hypothetical protein"
/protein_id="AA061874.1"
/db_xref="GI:34105520"
/translation="MPSBGRBQQLSGVLVASFMLKRLAPARALGLSPMCIGAVR
GIAMDALHGFASPSALSDDLVTYFADENLSPREALQRLASVHPDLPMEYTNQAA
VHLMFADRFGEVPSLESLALATWPEVATAGVALMNDSDLRILAPGLDLPAMT
VRNRPASRLSEIYRQVAOKRYQARMPRVTVLE"
complement(7321..8358)
/gene="ruvB"
/locus_tag="CV4215"
complement(7321..8358)
/gene="ruvB"
/locus_tag="CV4215"
complement(7321..8358)
/gene="ruvB"
/locus_tag="CV4215"
/notes="Identified by sequence similarity; putative; ORF
located using Blastx"
/codon_start=1
/translation="MIEDKLFGAAPERIVYPPORASBOEALERALPKLLDEVGQ
KRAEQLLEIFTEAKKRGALDHLVLFPGGLGKTLTAHIVAREGVNLRGTSPLVE
RAGDLALNTLEPHDLFIDRIHLSPVBEILYALPELOIDIMIGEPAASVVI
DLPPFLVATTRAGMLTNPLDRGIYARLEFYNAELTIVSSAGLNVQSDG
AFEVAKRSRGTPIANRLIRVRDYAEKSDGVVAADALMLDVPDGLDMVR
KLQATLEKPSGPGVGLDNVNAIGESTDTLEDVLEPPLIQGYLQRTPRGMATAQA
YLRGLPVDA"
8726..9358
/gene="ribB"
/locus_tag="CV4216"
8726..9358
/gene="ribB"
/locus_tag="CV4216"
/notes="Identified by sequence similarity; putative; ORF
located using Genemark/Blastx/COG0108"

Query Match      22.2%; Score 282.6; DB 1; Length 208524;
Best Local Similarity 56.2%; Pred. No. 3,7e-61;
Matches 572; Conservative 0; Mismatches 439; Indels 6; Gaps 2;

QY 15 CTCTCAATCAACAACCTGATGATTTGATCTTATCTCCGTGATGATGTTCTTAAAG 74
DB 130330 CTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 74
QY 75 GCAAGTGTCTCTCAACAGTGCACATCTTGTGGAGGCAATAGCAATGCCAATTGGAAG 134
DB 130270 GCGGTGCTGCCGACACCTGTGCGCAGATGGGCGGCCCATCTCATGCCCACTGAAG 130211
QY 135 CCTCTATCACTACCACTGCTGCTGTGTAGCATCCGGAGGCGCATATTGAATCTTAA 194
DB 130210 CCGCGGTGCTACCACTGCTGCTGTGTAGCATCCGGAGGCGCATATTGAATCTTAA 194
QY 195 CTTGTTGATAGTATTTCAACCTCTTATGACACTTATTTAGACAGATCAACAGCTCT 254
DB 130150 CCGGCGCGGACGCGCTTGAAGCGGCTGATGACCTTTATCTGACGATTAACAGCGCC 130091
QY 255 ATGGAATCAACATAGCAAGAGAGACCAAGTGTATTGAGGATGAAGTTGACCTGCT 314
DB 130090 GAGGAATTCGCAAGGCGCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 314
QY 315 GTTGCACGACAAATTTCTCAAGATGAGTGAATCTTTTGGGAAGTGTTTACAGTT 374
DB 130030 GGGGCGACACCAATTTCCGACCGGCTGACCGATAT---CGCCACGCGATGCGGCG 129974
QY 375 CTACAGAATGTTTGAATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 434
DB 129973 CTGAGAGCATGCGCAAGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 129914
QY 435 GAGGTGATGATGTTGATGAGAAAAGTATTCAATTGAACGTTCTTAACCGTTGCTG 494
DB 129913 GACATGACGTTGTCGACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 129854
QY 495 CAGAAATTTCCACATTTGAAGTCTGTATGAGACATTTACCATTTGATGCTGTTAAG 554

/product="conserved hypothetical protein"
/protein_id="AA061874.1"
/db_xref="GI:34105520"
/translation="MPSBGRBQQLSGVLVASFMLKRLAPARALGLSPMCIGAVR
GIAMDALHGFASPSALSDDLVTYFADENLSPREALQRLASVHPDLPMEYTNQAA
VHLMFADRFGEVPSLESLALATWPEVATAGVALMNDSDLRILAPGLDLPAMT
VRNRPASRLSEIYRQVAOKRYQARMPRVTVLE"
complement(7321..8358)
/gene="ruvB"
/locus_tag="CV4215"
complement(7321..8358)
/gene="ruvB"
/locus_tag="CV4215"
/notes="Identified by sequence similarity; putative; ORF
located using Blastx"
/codon_start=1
/translation="MIEDKLFGAAPERIVYPPORASBOEALERALPKLLDEVGQ
KRAEQLLEIFTEAKKRGALDHLVLFPGGLGKTLTAHIVAREGVNLRGTSPLVE
RAGDLALNTLEPHDLFIDRIHLSPVBEILYALPELOIDIMIGEPAASVVI
DLPPFLVATTRAGMLTNPLDRGIYARLEFYNAELTIVSSAGLNVQSDG
AFEVAKRSRGTPIANRLIRVRDYAEKSDGVVAADALMLDVPDGLDMVR
KLQATLEKPSGPGVGLDNVNAIGESTDTLEDVLEPPLIQGYLQRTPRGMATAQA
YLRGLPVDA"
8726..9358
/gene="ribB"
/locus_tag="CV4216"
8726..9358
/gene="ribB"
/locus_tag="CV4216"
/notes="Identified by sequence similarity; putative; ORF
located using Genemark/Blastx/COG0108"

Query Match      22.1%; Score 281; DB 6; Length 1071;
Best Local Similarity 56.1%; Pred. No. 1.3e-60;
Matches 571; Conservative 0; Mismatches 440; Indels 6; Gaps 2;

QY 15 CTCTCAATCAACAACCTGATGATTTGATCTTATCTCCGTGATGATGTTCTTAAAG 74
DB 43 CTCAATATTCGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 102
QY 75 GCAAGTGTCTCTCAACAGTGCACATCTTGTGGAGGCGCAATATCAATGCCAATTGGAAG 134
DB 103 ACCGTGTTCTTATACCAAGCTGTTATTTTGGGAGGCTATGCTAATGCCAATCTTGT 162
QY 135 CTCTCATATCACTACCTGCTGCTGTGATGACATACCGGAGGCGATATTGAATCTTAA 194
DB 163 CCGCGATCAACCAATTTGAAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 222

RESULT 11
AR375107 1071 bp DNA linear PAT 18-DEC-2003
LOCUS AR375107
DEFINITION Sequence 113 from patent US 6605709.
ACCESSION AR375107
VERSION AR375107.1 GI:40078289
KEYWORDS
SOURCE
ORGANISM
Unknown.
REFERENCE
1. (bases 1 to 1071)
AUTHORS
Breton, G.L.
TITLE
Nucleic acid and amino acid sequences relating to Proteus mirabilis
for diagnostics and therapeutics
JOURNAL
Patent: US 6605709-A 113 12-AUG-2003;
FEATURES
Location/Qualifiers
1..1071
ORIGIN
/organism="Unknown"
/mol_type="genomic DNA"

Query Match      22.1%; Score 281; DB 6; Length 1071;
Best Local Similarity 56.1%; Pred. No. 1.3e-60;
Matches 571; Conservative 0; Mismatches 440; Indels 6; Gaps 2;

QY 15 CTCTCAATCAACAACCTGATGATTTGATCTTATCTCCGTGATGATGTTCTTAAAG 74
DB 43 CTCAATATTCGCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 102
QY 75 GCAAGTGTCTCTCAACAGTGCACATCTTGTGGAGGCGCAATATCAATGCCAATTGGAAG 134
DB 103 ACCGTGTTCTTATACCAAGCTGTTATTTTGGGAGGCTATGCTAATGCCAATCTTGT 162
QY 135 CTCTCATATCACTACCTGCTGCTGTGATGACATACCGGAGGCGATATTGAATCTTAA 194
DB 163 CCGCGATCAACCAATTTGAAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 222
```


OY	195	CCCTGTGAAGTGAATTTCAACCCCTCTAATGACATTTAATTGACAGATACCAACAGCT	254
Db	223	CCTAGGCGCATTAATTCGAACCAATTATGACCTGCTATTATTAACGGATAGACCTCGCCA	282
OY	255	ATGAAATCAAACTAGCAGAGAGAGAGCCAGCTCGATTGTTGGGTGAAGTTACCTCT	314
Db	283	TCAGAAAGTGAACAAGGTTTTTTTCAAGGCGTATTACTGTTGTAAACCTTATCTGCC	342
OY	315	GGTGCACGACAAATTTCTCAAGATGAGTGACTGATCTTTTCGGAAAGTTTACCACTT	374
Db	343	AATCGACGACAAATTTCTAGCCATGAGTTCCTCATAT--CAATTAATCTATCTATT	399
OY	375	CTACAGAAAGTGTGAGCATATATGCTCTGCGGTTTCATGAGAGGTTACTAATCT	434
Db	400	CTTAGTGAATGAAGAAAATAGTATGCCATTGCTTATTCATGGGGAAGTTACTGCATCA	459
OY	435	GAGGTGACATGTTTGATAGAAAAGGTAATTCATGAAAGGTTCTAAGCCGTGTG	494
Db	460	GATATCGATATTTTGTATAGAGAAGCTGTTTTATTGTAACGATGCGTCCCGTTCG	519
OY	495	CAGAAATTTCCACATTTGAAGTGTGTATGAGCATGTTACCACATTGATGCTGTAA	554
Db	520	AAACATTTTCGTGCTTAAAAATGTTTTTAACACATACACACACAAAGACGACACG	579
OY	555	TTTGTGTAATCTTGACATGAAAGSATTGTCAGACATGTCACCCCAACATCTGTT	614
Db	580	TATGCTTAGAGGAAATGA---ATTCTTGTGTGAATATTAACACCAACACTATATG	636
OY	615	TTGAACAGGAATCTCTCTTCCAGAGGGGCTTCAACCGCATTAATTACTGCTTTCAC	674
Db	637	TTTAACCGTAATCATATGCTCTGTGCTGTGTGTAACCGCATTTGTACTGTTTACATAC	696
OY	675	CTCAAAAGAGATCCACAGGAGGCACTGTGTGAGCTGTAAACAAGTGAAGTAAAGA	734
Db	697	CTTAAGCGTAATGTACACCAAGGCAATTAAGCAAGCCGTTGCATCAGGCCATTGCG	756
OY	735	TTTTTCTTGGGACGTATAGTCTCTCATGATAGACGAAGAAAAGATGTTCTTGTGA	794
Db	757	TTCTCTTAGAGTACGTATTCCGCACTCATTTAACAACATCGAAGAAAGATCATCTGTGA	816
OY	795	TGTGCTGTATTATTAACATGACCTGTAGCTCTGTGATGATATGCAAGAGTGTGGAAG	854
Db	817	TGCGCTGGGTCTTTTAACGCCCCCAACCGCTTGAAGCTTATGCTAGAGGATTTAAAG	876
OY	855	GAAAAATGACCTGACAGAGCTTGAACATCTACTACTCAATGACACGATTTTATGG	914
Db	877	CTCAATGCCCTTCTCACTTTAAGCCTTCTGTATCACTCAATGGCCACGTTTTTATG	936
OY	915	CTTCTGAGAACACTCAAGATTAAGTTAGTAAAGACGCCATGGAAGTACCCGATCC	974
Db	937	TTACCGGTTAATGAAGGCACTATTACCTTAAACGAAAAATCCGTTACTGCACCTGC	996
OY	975	TTTTTCTTAGCATCAAGAGATTAATATCCACGTTTGTCTGGTGAATGCTGCAC	1031
Db	997	ATTATGAGTGTGATGAACATTAATCCCTCTTACGCAATGAAGATTAATTCATTGG	1053

RESULT 12	AE004420	11619 bp	DNA	linear	BCT 10-OCT-2003
LOCUS	AE004420				
DEFINITION	Vibrio cholerae O1 biovar eltor str. N16961 chromosome II, section 77 of 93 of the complete chromosome.				
ACCESSION	AE004420	AE003853			
VERSION	AE004420.1	GI:9658361			
KEYWORDS					
SOURCE					
ORGANISM	Vibrio cholerae O1 biovar eltor str. N16961 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.				
REFERENCE	1 (bases 1 to 11619) Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Hail, D.H., Hickey, E.K., Peterson, J.D., Unayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,				
AUTHORS					

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*
Nature 406 (6795), 477-483 (2000)

20406833
10952301

2 (bases 1 to 11619)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickley, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Teitelman, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Basse, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Uitterback, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.

TITLE
JOURNAL
FEATURES

Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers

gene
215. .1267
/gene="VCA0925"
215. .1267
/gene="VCA0925"
/note="similar to PID:1653627; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydroorotase"
/protein_id="AAF96822.1"
/db_xref="GI:9658362"
/translation="MHSWHSMTLTITRPDDMHVHLRDDGVLDADTVRDISRNGR
LIMAPTPPVVTTTENALVYRERIMAAQQAIFEPIMALYLDNTSPEETIRAKASGKRL
VAAKYIPGATITSDSGVTSAKNIYPRVIAQAEVGMALVGEVTTHEVDIFDEKRT
LDYVATPVPNPOLKI VLEHITTTADATVFOQAGDVAATITAHNLFFNNHLYVGG
ITRPHVYCIPIKRAITHQALVALAATSGSKFFELGDSAPAHKGRKEAACGAGSYTA
AALETXAVPEKECKLENNLEAPASFNQSPDFGLPRNQSTVTLTQAMVVASMPGSI
LVPYRAGENIENTYK"
complement(1330. .2427)
/gene="VCA0926"
complement(1330. .2427)
/gene="VCA0926"
/note="similar to GP:2906236; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, ArcC/XylS family"
/protein_id="AAF96823.1"
/db_xref="GI:9658363"
/translation="MRNTOVQGLPFTFGSKDMLMIVPFPVSLMLLLATLYRLRL
EAKAKACIEIVCAATTAIVVGRMTWDVSLRMAPDILASILPVAMCTPRTFKHGK
SCKYKLSPTIIVLYSLAOPFWSLPDIBILITAFVIGILNRYSDQDAIVYAFV
HNEVYKRAKNIIGMMLFSALIDSAMSDPITNQGRELYITVGHVILIPYLAIV
LVGTSITSSDEVAANLEITTSHEVKEKTLIDEGMPEARAEIVALEPRORKEKHYIDP
WLSKTSRLGIPAKOISAAVNOVKNIKISKLINEYRIDHMLALTQSDDSITQIFPM
SGFOITSNRNEFRSRTVGTMTPEYRKQRHN"
2538. .2888
/gene="VCA0927"
2538. .2888
/gene="VCA0927"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF96824.1"

QY 435 GAGGTGACATGTTTGTAGAGAAAAGTATTCATTGAAACGGTTCAGACCGTGTG 494
DB 662 GAAGTGCACATTTTGTGACCGTGAAAAGACCTCTCTACACACCGTGTGCGCATG 721
QY 495 CAGAAATTTTCACATGAGAGTGTGATGAGACATGTTTCCACATTTGATGCTGTTAAG 554
DB 722 AATGATTTCCCGACACTGAAATTTGTCTTGAACATCAACACCGCGCATGCAATGACC 781
QY 555 TTTGTGTAATCTTGACATGAGGATTTGTTGACGAACTGTCAACCCCAACATCTTGT 614
DB 782 TTTGTGCA---CAAGCAGGCGATTAAGTTGCGGCAACGATTCGCGCACATTTGCTG 838
QY 615 TTGAAACAGGAATCTCTCTTCCAAAGGGGCTTACACCGCATATTTACCTCTTCCAGTC 674
DB 839 TTTAAACCGTAACCAACATGCTGTGGGTGATTCGCCACATTTTACTGTTTGCATC 898
QY 675 CTGAAAGAGAGATTCACAGGAGGACCTGTGTGACGTGTAACAGTGAAGTAAAGA 734
DB 899 CTAAAGGCTGACATCCACACGCGCTTGTGACAGGCACTCTGCGACGAAAGAA 958
QY 735 TTTTCTTGGGACTGATAGTCTCTCATGATAGACGAAAGAAAGTCTTGTGGA 794
DB 959 TTTCTCTAGGTACGACCTGCGCCGACGCAAGGCGCAAGAAAGCGCTTGTGCGC 1018
QY 795 TGTGCTGATTTTACATGCACTGTGACCTTGTGATATGCGAAAGGTGTTGAAG 854
DB 1019 TGGCAGAGTCTTACACAGCCACGACCTTGTGATGATGCGAAAGTGTGAAAGAA 1078
QY 855 GAAATGACATCGCAAGCTTGAAGCACTGACCTTCAATGAGACCAAGATTTTATGCG 914
DB 1079 GAAGCTAAGCTAGAAATCTGGAAGCGTTGCTTGAATGAGCGCTGATTTTCTAGCGC 1138
QY 915 CTCTCTAGGAACTCAAGATTAAGTTAGTTAGTAAGACGCCATGAAGTACCGAAATCC 974
DB 1139 CTACCTGCAATCAAGAGACCGCTCACGCTCAACGAAGCGCTGCGCATGCGAAGAC 1198
QY 975 TTTTCTTATGATCAGAGATATTTATTCCTGTTTGTGTGTAATGCTGACTGG 1031
DB 1199 ATGCGGTTTGGCAGCAGATATGCTGTGCAATCCGCGCGGTGAAGATGAGTGG 1255

RESULT 13
BX572091 349082 bp DNA linear BCT 10-0UN-2004
LOCUS Prochlorococcus marinus MED4 complete genome; segment 2/5.
DEFINITION BX572091 BX548174
ACCESSION BX572091 GI:33639658
VERSION
KEYWORDS complete genome.
SOURCE Prochlorococcus marinus subsp. pastoris str. CCMPI986
ORGANISM (Prochlorococcus marinus MED4)
Bacteria; Cyanobacteria; Prochlorales; Prochlorococcaceae;
Prochlorococcus.
1 (bases 1 to 349082)
Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P.,
Ahlgren, N.A., Arellano, A., Coleman, M., Hauser, L., Hees, W.R.,
Johnson, Z.I., Land, M., Lindell, D., Post, A.F., Regala, W., Shah, M.,
Shaw, S.L., Steglich, C., Sullivan, M.B., Ting, C.S., Tolonen, A.,
Webb, E.A., Zinser, E.R. and Chisholm, S.W.
Genome divergence in two Prochlorococcus ecotypes reflects oceanic
niche differentiation
Nature 424 (6952), 1042-1047 (2003)

TITLE
JOURNAL MEDLINE
22825698
12917642
PUBMED
2 (bases 1 to 349082)
Larimer, F. and Rocap, G.
Prochlorococcus genome consortium
Direct Submission
Submitted (03-JUN-2003) Submitted on behalf of the Prochlorococcus
genome consortium, the DOE Joint Genome Institute, Production
Genomics Facility, 2800 Mitchell Drive, Walnut Creek, CA 94598,
USA, and the Genome Analysis Group, Oak Ridge National Laboratory,
1060 Commerce Park Drive, Oak Ridge, TN 37831, USA.

FEATURES
source
larimerf@ornl.gov
Location/Qualifiers
1..349082=Prochlorococcus marinus subsp. pastoris str.
/organism="Prochlorococcus marinus subsp. pastoris str."
/mol_type="genomic DNA"
/strain="MED4"
/db_xref="taxon:59919"
/note="synonym: Prochlorococcus marinus MED4, strain
coliditicy: CCMPI378 = MED4"
2503..2835
/locus_tag="PMM0369"
2503..2835
/locus_tag="PMM0369"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="possible Ribosomal protein S14p/S29e"
/protein_id="CAE18828.1"
/db_xref="GI:33639659"
/translation="MLQKRAKLETFDRIREADIVIEGLAFSSITSYKRYTTSR
RQARRIADVAVKIVRTDIESDIDRKEMTSKLVCLLENTYHLMKEKQYSAVATNARV
LMRLIRLE"
3283..4968
/locus_tag="PMM0370"
3283..4968
/locus_tag="PMM0370"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative cyanate ABC transporter, substrate
binding protein"
/protein_id="CAE18829.1"
/db_xref="GI:33639660"
/translation="MLNPVNSYTKKYNFKCGPSADNSDQFQDONSLSNLLNPANP
HGLYCAVECPRIKNDHENSMDPDPVPIISDFSPMGILIKQAYVADAISTENR
DLFVKNASNDGPDKKELRELAECAGLPAPAPAGPQAGFANTRAAISPAEGR
PLSLAAGALITVACSONKPKPDPDPININLEKOTLRVGFPIICASPIIAA
EPWGFYKRFMDKVVVKNPSPKAVRSDAINGEDAYTMILPMPISMTLIGCTSPFSK
LASIENTNGAIVVANKHKKVKGADMKQPIFGVPPSYMHLILARYILAKGVDPD
KDVQIRVPPPPDIQLVADIDAYLMDPPFNQRAVEDAGFIHLTKELMPGHPCA
FAAGEPWINSPEFTRALNSIIDAAYVSTPSNRKEVAKAISGRGPLNDPTEVEAV
LTGKFEDGLKTONVPRIDPKPYMOSFSHMISOLIRMDLGGAVAIAGADFNPS
ASIFLIDBAQALREKGNPPTARFREVALAYDEFDPSNP1KVKQDKIDGPF"
4999..5781
/locus_tag="PMM0371"
4999..5781
/locus_tag="PMM0371"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative cyanate ABC transporter"
/protein_id="CAE18830.1"
/db_xref="GI:33639661"
/translation="MNKVISKRLTFVVTLSLFIITIMOIAGOGLLVKNPSPGPKT
INALIMVSPFPFNGNDIGCNLLISLRUYIGSLAVLVIPILGVGKMSKLAQ
SSLNPVQLKPVSLMDPLVGLVFFNSSEITGTFVITPITSIMPTLINTAGVASINP
DYLRVSKLSIGNSKSIITIRVILPALIMNIIAGMISMGTAALVVAEMLMGTGIGIF
IMENMNLSENIFVAILIIGTFIIDQFPGFIQDKRFVYSI"
4999..5088
/locus_tag="PMM0371"
/note="Signal predicted by SignalP 2.0 HMM (Signal peptide
probability 0.892) with cleavage site probability 0.817 at
residue 30"
order(5023..5082,5212..5280,5404..5472,5590..5658,
5677..5745)
/locus_tag="PMM0371"
/note="5 probable transmembrane helices predicted by
TMHMM2.0"
5798..6652
/locus_tag="PMM0372"
5798..6652
/locus_tag="PMM0372"
CDS
gene
misc_feature

gene
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative cyanate ABC transporter"
/protein_id="CAE18831.1"
/db_xref="GI:33639662"
/translation="MQRKSNKRSSTHLEIAGIKGIFFEEKFPDPYAKHKSFEAFI
KQNIILKENTPVSITGPGSGKSTLNTLAGLAPTSGKIIIDGLAIKGGPRGVI
FQNYILMPPTTIIQIEIRLMDLWPEPKTKPIDIERAIIYILRWGEEKSNFPELSG
GKQKRYALRARSINPNNLIMDEPFGLDALDLSYLOEILSIWRKNVTVVLTTHI
EHALLSKIIILMSGPSATISEEISVDLRPRDRQNI EQDPKYLKIMRLKHLRE
TVAVEEVV"
6685..7128
/gene="cynS"
/locus_tag="PMN0373"
6685..7128
/gene="cynS"
/locus_tag="PMN0373"
/EC_number="4.2.1.104"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="Cyanate lyase"
/protein_id="CAE18832.1"
/db_xref="GI:33639663"
/translation="MTLPSTLTKLLKAKKKNLSPDLGNLIRDEWVASLFYGOAT
VSEEEATSLIADLITSDLKEDLSTPVGGLDVIPTDPLITRFELMQVGLPWKD
VIQKRFPGDIMSIIIDSEIVDEVDPEKGDRLVLYKCGKFLPYKKW"
complement (7213..7404)
/locus_tag="PMN0374"
complement (7213..7404)
/locus_tag="PMN0374"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="meta/Hcf106 family"
/protein_id="CAE18833.1"
/db_xref="GI:33639664"
/translation="WNITFGISLPEIIVILILGLIFGPKRLPOLGKTIQKTLKQIQA
SKFESEIKIKTLQNLEND"
complement (7336..7392)
/locus_tag="PMN0374"
/locus_tag="PMN0374"
/note="1 probable transmembrane helix predicted by
TMHMM2.0"
7504..7827
/locus_tag="PMN0375"
7504..7827
/locus_tag="PMN0375"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="possible Type II intron maturase"
/protein_id="CAE18834.1"
/db_xref="GI:33639665"
/translation="MERSMDPYKYFKPEDFNLFEIPLHKEFOKEIIDLGLYQNEIN
QFNVPGQFCVWNGVIKIYHDLPIFLMDKKPRDNINIEKMLDSYINLSTIDIKWVES
SIKII"
complement (7843..8211)
/locus_tag="PMN0376"
complement (7843..8211)
/locus_tag="PMN0376"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical"
/protein_id="CAE18835.1"
/db_xref="GI:33639666"
/translation="MKLSLSAVLFLFTLSFAQEKLNVTSDSQIOSIKGNPAIG
NVYIKNNINFEASNNKLTYDDAKTLKLVGNVFNVLSEGLSIKSGIDELTFTD
SGLFKRNISNKGKRVKTKF"
complement (8152..8211)
/locus_tag="PMN0376"
/note="Signal predicted by Signalp 2.0 HMM (signal peptide

[illegible]

Qy	672	GTCTCAAAAGAGATCCACAGGAGGACCTGTGTCTACCTGTAACAAGTGGAAATAA	731	/translation="MIYKOVHADRSPSLFTSKFVFPGATFFCCLLMGSAVPAIKNGYV LNIISPDILPSKMFVAGVGLVLLFALASGAIGRLKGOYQOQAIILGLTQTA LOVPEYGLATTTGKSGISIMNATGTFEVLVAHYLYKXDRLLTLNKICGALGPGVM VNVGSDLMDFNPSLIGBGSVLLAAFIILSAITYGRISQMDTTVMTEYQALIGLV LTLGYLEVFFGLQ"
Db	187911	GTCCAAAGAGAGAAAACAATGGAATTCCTTAAAGAAAGCTGCTACTTAGGAAAAAA	187970	
Qy	732	AGATTTTTCCTGGAGCTAGTAGTGTCTCATGATAGACGAGAAAAAGATGTTCTTGT	791	/gene=".1126 965. .1126 /gene="y1735" 965. .1126 /gene="y1735" /function="unknown" /codon_start=1 /transl_table=11 /product="hypothetical" /protein_id="AA085303.1" /db_xref="GI:21958539" /translation="MSGAVASGFDKSKSTVAHFLLCSTAGVTSFVPAKIKTRIVLFTY LRIDLRAIH"
Db	187971	TGTTTTTCATAGGACGAGATTCGTCTCTCACTTAAGACAGTGGAAAG---CTTTTGT	188027	
Qy	792	GGATGTCTGTGATTTTACAATGACACCTGTAGCCCTGTTCAGTATATGCGAAGTGTGAA	851	
Db	188028	GTTGTGACAGGCAATTTTAAATTTCTCCAGTAGCAATAGAAAGCTATTTAAAAAGTTTGTAG	188087	
Qy	852	AAGAAAAATGACATCGACACAGCTGGAAGCATTTCACTAGCTTCATGACGACGATTTTAT	911	
Db	188088	GAGGAAAAATGACCTTAATTCATTTGAAAAAATTTGCCAGCTGAATGGCCCTAAATTTTAT	188147	
Qy	912	GGGCTCTCTAGCAACAACCTCAAGATTAAGTTAGTGAAGCCGACGGAAGGTACCCGAA	971	
Db	188148	AATTCACCTATTAATTAACGAAAAAATTAAGATTTGATATCAAGGTCAATTAATTCAGAA	188207	
Qy	972	TCCTTT 977		
Db	188208	TTTATT 188213		
RESULT 14				
LOCUS	AE013777	10421 bp	DNA	linear
DEFINITION	Yersinia pestis KIM section 177 of 415 of the complete genome.			
ACCESSION	AE013777	AE009952		
VERSION	AE013777.1	GI:21958537		
KEYWORDS				
SOURCE	Yersinia pestis KIM			
ORGANISM	Yersinia pestis KIM			
REFERENCE	1 (bases 1 to 10421)			
AUTHORS	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Lies, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.			
TITLE	Genome Sequence of Yersinia pestis KIM			
JOURNAL	J. Bacteriol. 184 (16), 4601-4611 (2002)			
PUBMED	12142430			
REFERENCE	2 (bases 1 to 10421)			
AUTHORS	Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F., Lies, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C., Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V., Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S., Blattner, F. R. and Perry, R. D.			
TITLE	Submitted Submission			
JOURNAL	Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445			
PUBMED	Henry Mall, Madison, WI 53706, USA			
REFERENCE	Location/Qualifiers			
AUTHORS	1. .10421			
TITLE	/organism="Yersinia pestis KIM"			
JOURNAL	/mol_type="genomic DNA"			
PUBMED	/strain="KIM"			
REFERENCE	/db_xref="taxon:187410"			
AUTHORS	230. .928			
TITLE	/gene="y1734"			
JOURNAL	230. .928			
PUBMED	/function="putative membrane"			
REFERENCE	/notes="residues 10 to 226 of 232 are 70.96 pct identical			
AUTHORS	to residues 5 to 220 of 308 from Genpept : >gb AAJ2325.1			
TITLE	(AE008906) putative permease [Salmonella typhimurium LT2]"			
JOURNAL	/codon_start=1			
PUBMED	/transl_table=11			
REFERENCE	/product="putative membrane protein"			
AUTHORS	/protein_id="AA085302.1"			
TITLE	/db_xref="GI:21958538"			
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				

to residues 1 to 340 of 344 from E. coli K12 : B1601;
residues 10 to 349 of 361 are 62.05 pct identical to
(JAB009764) putative membrane transport protein [salmonell
lythiumum LT2]"
/codon_start=1
/transl_table=1
/product="putative transport protein"
/protein_id="AA065307.1"
/db_xref="GI:21958543"
/translation="MAVRIQGLITTPFVSNQGLRVAMLAFLVVMAGIKRASPITYV
PRLIAFLAVINLNPVLKLERIKIPRLIAVILLVITIIILAVLFLGRLSSINPASP
LPQIKMVELEKRRDQEPALFNIEFSDVMKCYVSPSNAMPVRLISLHGAMTST
FLILVTVMFLFEVERMPYKVALFDNPEKGNAILKRAANGVTHYIVLTKTISTATGG
VLMFLAAMGVAFALMGLAFVANYIPNIGSVLAIIPPIOTALFNGPDAFAVTGGG
YLILMIGNIIIDPRMGRGLGSLTVVFLSLFMGLIGPVGMLLSVPLTIIAKIVL
ELTPGVGFVAILDQGRAPKVEEQAL"
complement(4368..4481)
/gene="Y1740"
complement(4368..4481)
/gene="Y1740"
/note="residues 4 to 37 of 37 are 47.05 pct identical to
residues 12 to 45 of 46 from E. coli K12 : B1058"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA085308.1"
/db_xref="GI:21958544"
/translation="MFKLLSMVYNVREHLLLYICSLVLSGLDIPFFY"
complement(4825..5466)
/gene="Y1741"
complement(4825..5466)
/gene="Y1741"
/function="unknown"
/note="residues 10 to 199 of 213 are 28.71 pct identical
to residues 2 to 187 of 199 from GenPept : >gblAAC46458.1
(U2172) entericidin R [Citrobacter freundii]"
/codon_start=1
/transl_table=1
/product="hypothetical"
/protein_id="AA085309.1"
/db_xref="GI:21958545"
/translation="MLSDQVMHKKILIIDCEFYTSQVNTLYNSVSIKMDVPTVE
EALIAIQEQPDIIVNLVYCRGQHCPPLAQPIHYCTEAKYIVLDAVSFESTPI
PLTBSVSMPIARTKSIVSLIAKHLPNLLRRIQFSLERLGIQNOOASLFSQEHKVM
CYMTMEPYRIARLKINSGTSVYSHRHLTEKIKVKNRLCELTINVKIY"
complement(5706..6821)
/gene="sola"
/note="synonym: Y1742"
complement(5706..6821)
/gene="sola"
/function="putative"
/note="residues 1 to 371 of 371 are 64.24 pct identical to
residues 1 to 372 of 372 from E. coli K12 : B1059"
/codon_start=1
/transl_table=1
/product="arcsine oxidase-like protein"
/protein_id="AA085310.1"
/db_xref="GI:21958546"
/translation="MDYDLIVISSGVSAGAYVASOGLANVLMIDSAMPHQASHH
GSTRIMRAVSEGEKIVPLVRAQALMDLQATEKLFQAGVYINIGPNSIKATVE
QRSADQDLPEVLTNSTOIREKMPVPTVDNVIAYFEOSQGLSSELAIVETLIKATVE
AGCGILFNCPEVTALIESHQADVVTIDTGGYKVVYAVVWEELPTLPVTPKRV
FSMHQGLDESEANHPAPFVEMPDNIIVYGGPANDALKIGKHHGQGLIESAQKRP
FGRAEDQTEVFSFLRHLFPLRGVGCIRGACGSDYSDPEDPTIDTLFEDERVVVSGL
SGHGFALDTEVFTALFADKRSPIDISATSLAKR"
7111..8425
/note="insertion element"
/insertion_seq="IS285"
7182..8390
/gene="Y1743"
7182..8390
/gene="Y1743"

```
function="IS and transposon related functions"
/notes="IS285: residues 1 to 402 of 402 are 100.00 pct
identical to AF053197.1 (AF053197) transposase [Yersinia
pestis]"
/codon_start=1
/transl_table=1
/product="putative transposase"
/protein_id="AA065311.1"
```

Query Match	21.8%	Score 277.2;	DB 1;	Length 10421;
Best Local Similarity	55.9%	Pred. No. 1.1e-59;		
Matches 569; Conservative	0.0	Matches 443		

QY 15 CTCTCAATCACAAACCTGATGATGGCATCTTCATCTCCGATAGTGATGTTTAAAG 74
 Db 9210 CTAATAATTCGCGCCGGATGATCGGCACATTCACTACGTAATGATGAATGCTCAGT 9265
 QY 75 GCAGTGTCTCTCACACTGCAATCACTACTTTGGAGGGCAATGATGATGCCAAATTTGAAG 134
 Db 9270 ACCGTGTGCTTAACTCTCGAAGTATTCGCCCGGCTAATTGTATGCAAAATCTAAC 9329
 QY 135 CCTCTATCACTACCACTGCTGTCTGTAGATATCCGGAGGCAATTTGAATCTTTA 194
 Db 9330 CAGGCATATTCAAACGGTTCGCACTGTAATTCCTTAACGGAGCGTAATTTTGAACCGGT 9389
 QY 195 CCGTGTATAGTATTTCAACCCCTCTTATGACACTTTATTTGACAGATACCAACAGTCT 254
 Db 9390 CCTGGGGCCATTAATTCACCCCGTGTATGACGTGTACTGTAATTAAGCTTGATGCT 9449
 QY 255 ATGAAATCAATCTAGCAGAGAGAGCCAGGTCGATTTTGGGGTGAAGTTGACCTGCT 314
 Db 9450 AAGAGTTGACACCGGGTTTGGACAAAGCGTTTACCGGGGCAACGTGATCCCGCC 9509
 QY 315 GGTGCCACGCAAAATTTCTCAGATGAGTGACTGATCTTTTGGGAGTGTTTACCACT 374
 Db 9510 AATGCCACCAACACTCCACTCACGCTGATCTGACAT---CCGGCAATTAACCGTTG 9566
 QY 375 CTACAAGAATGTTGAGCATATATATGCTGTGCTGTATAGAGAGTTACTAATCT 434
 Db 9567 TTGTGAACAAATCAGAAAGATAGGCAATGCCCTGTTTATCAAGTAGAGTAAAGATGG 9626
 QY 435 GAGGTGGACATGTTTGAATAGAGAAAGATTTCAATTGAAACGGTCTAAGACGGTGTG 494
 Db 9627 GCGGTGACATCTTTGATCGTGAAGCCCGTTTATGACAAATTTTGAAGGCCATTCGC 9686
 QY 495 CAGAAATTTCCAAATGGAAGTGTGATGAGACATGTTACCACTTGAATGCTGTTAAG 554
 Db 9687 CAAAAGTTTCCGAACTTAAATTTGTCTTTGAGCATATACGACCAAAATGCGCAGAT 9746
 QY 555 TTTGTTGAATCTTGCACTGAAGGATTTGTTGACAACTGTCAACCCCAACATCTGTT 614
 Db 9747 TATGTGCTGGAGGCA---ATCGTTTCTTGGGGCAACGTCACGCGCAACACTTGATG 9803
 QY 615 TTGAACAGGAATTTCTCTCTCAAGGGGGCTTACCAACCGCATTAATTACTGCTTCACT 674
 Db 9804 TTAAACCGCATATCAATGCTGTGTAAGGGGATATTCGCCCTCACTGTTTGTGCTCGCAATA 9863
 QY 675 CTCAAAAAGAGATCCACAGGGAGGCACTTGTGATGCTGTAAACAATGGAAGTAAAGA 734
 Db 9864 TTGAAGGCGACACCATCATGACAGCATTTGCCGCGACGCTGCCAGTGCTGTGATGCG 9923
 QY 735 TTTTCTTGGAGCTGATGCTGCTCTCATGATACGAAAGAAAGAGTGTCTTTGTGA 794
 Db 9924 TTTCTCTTGGAGCCGATTCAGCTCCCAATGCAAAACGTGAAGAGTCAATCTTTGGCGC 9983
 QY 795 TGTGCTGTGATTTTCAATGCACCTGTAGCTGTGTCAATATGCGAAGGTGTTGAAAG 854
 Db 9984 TGTGCGGTGTATTCAAACGCCCAACGCGCATTTGCTGTATGCTTCCGTGTTTGAAGAA 10043
 QY 855 GAAATATGCACTGCAACAGCTTGAAGACATTCATGACTTCAATGAGCAACAGATTTTATGG 914
 Db 10044 CTGAATGCAATTCGAACATCTGGAAGCGTTTGGCGCTTAAATATGGCCACACATTTTATGGC 10103

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: February 12, 2005, 07:04:55 ; Search time 80 Seconds
(without alignments)
1413.190 Million cell updates/sec

Title: US-10-070-277-2

Perfect score: 1812
Sequence: 1 MELSTQPDMDHLRQGDV.....SGDIIMFAGXGMDWLPAPL 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376875

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US10A_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppaa/US10D_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppaa/US11_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*
20: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1528	84.3	345	US-10-451-554-7	Sequence 7, Appl1
2	1458	80.5	359	US-10-437-963-189313	Sequence 189313,
3	1454	80.2	379	US-10-424-599-149491	Sequence 149491,
4	1067.5	58.9	364	US-10-282-122A-48169	Sequence 48169, A
5	1058	58.4	352	US-10-282-122A-50426	Sequence 50426, A
6	1039.5	57.4	350	US-10-282-122A-77674	Sequence 77674, A
7	1008.5	55.7	354	US-10-282-122A-5186	Sequence 5186, A
8	983	54.2	344	US-10-282-122A-65811	Sequence 65811, A
9	979	54.0	344	US-10-282-122A-65238	Sequence 65238, A
10	978	54.0	348	US-10-282-122A-59728	Sequence 59728, A
11	975	53.8	348	US-10-282-122A-66489	Sequence 66489, A
12	971	53.6	348	US-10-282-122A-43220	Sequence 43220, A
13	962	53.1	348	US-10-282-122A-67595	Sequence 67595, A

14	959	52.9	347	US-10-282-122A-69927	Sequence 69927, A
15	956	52.8	348	US-10-282-122A-78404	Sequence 78404, A
16	952.5	52.6	344	US-10-282-122A-44700	Sequence 44700, A
17	949	52.4	348	US-10-282-122A-75878	Sequence 75878, A
18	943	52.0	354	US-10-282-122A-63114	Sequence 63114, A
19	934	51.5	348	US-10-282-122A-72752	Sequence 72752, A
20	837	46.2	260	US-10-767-701-31604	Sequence 31604, A
21	569	31.4	335	US-10-282-122A-54267	Sequence 54267, A
22	541	29.9	339	US-10-282-122A-58744	Sequence 58744, A
23	539.5	29.8	190	US-10-282-122A-55550	Sequence 55550, A
24	527	29.1	339	US-10-335-977-6923	Sequence 6923, Ap
25	527	29.1	371	US-10-335-977-51620	Sequence 6923, Ap
26	420	23.2	270	US-10-767-701-51620	Sequence 51620, A
27	402	22.2	136	US-10-282-122A-68916	Sequence 68916, A
28	355	20.1	87	US-10-424-599-14940	Sequence 14940, A
29	318	17.5	86	US-10-767-701-37668	Sequence 37668, A
30	262	14.5	82	US-10-424-599-250183	Sequence 250183, A
31	209.5	11.6	146	US-10-335-977-6922	Sequence 6922, Ap
32	144	7.9	396	US-10-282-122A-52409	Sequence 52409, A
33	131.5	7.3	426	US-10-282-122A-60843	Sequence 60843, A
34	128	7.1	424	US-10-282-122A-52160	Sequence 52160, A
35	121.5	6.7	427	US-09-815-242-10858	Sequence 10858, A
36	121.5	6.7	427	US-10-282-122A-57073	Sequence 57073, A
37	119.5	6.6	426	US-10-282-122A-57626	Sequence 57626, A
38	117	6.5	421	US-10-282-122A-52993	Sequence 52993, A
39	115.5	6.4	428	US-10-156-761-14392	Sequence 14392, A
40	111	6.1	2198	US-10-369-493-5496	Sequence 5496, Ap
41	110	6.1	422	US-10-282-122A-72383	Sequence 72383, A
42	108	6.0	425	US-09-815-242-5884	Sequence 5884, Ap
43	107.5	5.9	425	US-10-282-122A-70902	Sequence 70902, A
44	106.5	5.9	432	US-10-472-928-2332	Sequence 2332, Ap
45	105.5	5.8	323	US-10-424-599-155522	Sequence 155522, A

ALIGNMENTS

```
RESULT 1
US-10-451-554-7
; Sequence 7, Application US/10451554
; Publication No. US20040111769A1
; GENERAL INFORMATION:
; APPLICANT: Rodolphe Arthur Kanhouou
; APPLICANT: Ramon Serrano Salom
; APPLICANT: Roque Ros Palau
; TITLE OR INVENTION: Sugar Beet genes involved in stress tolerance
; FILE REFERENCE: 1187-24
; CURRENT APPLICATION NUMBER: US/10/451,554
; PRIOR APPLICATION NUMBER: PCT/EP01/15093
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: EP 00870319.1
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 60/271,656
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Beta vulgaris
US-10-451-554-7
Query Match 84.3%; Score 1528; DB 16; Length 345;
Beet Local Similarity 82.0%; Pred. No. 5.6e-147;
Matches 282; Conservative 34; Mismatches 28; Indels 0; Gaps 0;
QY 1 MELSTQPDMDHLRQGDVYKAVSHHFGAIVMPNKKPITTTAAVAAREAILK 60
Db 1 MELSTQPDMDHLRQGDVYKAVSHHFGAIVMPNKKPITTTAAVAAREAILK 60
QY 61 SLVDSFNPPLMTLYLDTTSPMEIKLARISQVFGVLYPAGATTSQDQVTLDFKCL 120
```

Db 61 VLPPDSDFNPLMTLYLTDTPSPNEIKLARKSEVYVAKLYPAGATTNSODGVTDLGGKCL 120
Qy 121 PVLQEMVENHMPPLVHGEVTNPEDVMDREKVFIEITVLRPLVQKFPOLKVMMEHTTTIDA 180
Db 121 PVLQEMVENHMPPLVHGEVTNPEDVMDREKVFIEITVLRPLVQKFPOLKVMMEHTTTIDA 180
Qy 161 VKFVESCCEGVAATVTPQHLVLRNSLFOGGLQPHNYCLPVLKREIHEALVSATSGS 240
Db 161 VKFVESCCEGVAATVTPQHLVLRNSLFOGGLQPHNYCLPVLKREIHEALVSATSGS 240
Qy 241 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 300
Db 241 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 300
Qy 301 YGLPRNNSKIKLSTKTPKVPESFVSAGDIIPMFAGEMLDWLPAP 344
Db 301 YGLPRNNSKIKLSTKTPKVPESFVSAGDIIPMFAGEMLDWLPAP 344

RESULT 2

US-10-437-963-189313
Sequence 189313, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Bardazuk, Brad
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 189313
LENGTH: 359
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_85834C.1.pap
US-10-437-963-189313

Query Match 80.5%; Score 1458; DB 16; Length 359;
Best Local Similarity 78.1%; Pred. No. 8.3e-140;
Matches 268; Conservative 39; Mismatches 36; Indels 0; Gaps 0;

Qy 2 ELSTTODDWHHLIRDDGDLKAVVSHSAHFGRAIVMPNKKPITTTAAVAAYREAILKS 61
Db 12 ELSTTODDWHHLIRDDGDLKAVVSHSAHFGRAIVMPNKKPITTTAAVAAYREAILKS 61
Qy 62 LPPVDSDFNPLMTLYLTDTPSPNEIKLARKSOVVGKLYPAGATTNSODGVTDLGGKCL 121
Db 72 LPPVDSDFNPLMTLYLTDTPSPNEIKLARKSOVVGKLYPAGATTNSODGVTDLGGKCL 121
Qy 122 VLQEMVENHMPPLVHGEVTNPEDVMDREKVFIEITVLRPLVQKFPOLKVMMEHTTTIDA 181
Db 122 VLQEMVENHMPPLVHGEVTNPEDVMDREKVFIEITVLRPLVQKFPOLKVMMEHTTTIDA 181
Qy 132 VLEEMARQEMPLVHGEVTNPEDVMDREKVFIEITVLRPLVQKFPOLKVMMEHTTTIDA 191
Db 132 VLEEMARQEMPLVHGEVTNPEDVMDREKVFIEITVLRPLVQKFPOLKVMMEHTTTIDA 191
Qy 182 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 300
Db 182 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 300
Qy 242 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 301
Db 242 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 301
Qy 301 YGLPRNNSKIKLSTKTPKVPESFVSAGDIIPMFAGEMLDWLPAP 344
Db 301 YGLPRNNSKIKLSTKTPKVPESFVSAGDIIPMFAGEMLDWLPAP 344

RESULT 3

US-10-424-599-149491
Sequence 149491, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 149491
LENGTH: 379
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(379)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_106012C.1.pap
US-10-424-599-149491

Query Match 80.2%; Score 1454; DB 15; Length 379;
Best Local Similarity 78.0%; Pred. No. 2.3e-139;
Matches 269; Conservative 36; Mismatches 40; Indels 0; Gaps 0;

Qy 1 MELSTTODDWHHLIRDDGDLKAVVSHSAHFGRAIVMPNKKPITTTAAVAAYREAILKS 60
Db 34 MELSTTODDWHHLIRDDGDLKAVVSHSAHFGRAIVMPNKKPITTTAAVAAYREAILKS 60
Qy 61 LPPVDSDFNPLMTLYLTDTPSPNEIKLARKSOVVGKLYPAGATTNSODGVTDLGGKCL 120
Db 94 LPPVDSDFNPLMTLYLTDTPSPNEIKLARKSOVVGKLYPAGATTNSODGVTDLGGKCL 120
Qy 121 VLEEMARQEMPLVHGEVTNPEDVMDREKVFIEITVLRPLVQKFPOLKVMMEHTTTIDA 180
Db 154 VLEEMARQEMPLVHGEVTNPEDVMDREKVFIEITVLRPLVQKFPOLKVMMEHTTTIDA 180
Qy 181 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 300
Db 214 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 300
Qy 241 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 301
Db 274 KRFFLGTSAPHDRRREKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGPDF 301
Qy 301 YGLPRNNSKIKLSTKTPKVPESFVSAGDIIPMFAGEMLDWLPAP 345
Db 334 YGLPRNNSKIKLSTKTPKVPESFVSAGDIIPMFAGEMLDWLPAP 345

RESULT 4

US-10-282-122A-48169
Sequence 48169, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyckind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert

```

; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48169
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-48169

```

```

Query Match      58.9%; Score 1067.5; DB 15; Length 364;
Best Local Similarity 60.4%; Pred. No. 6.3e-100;
Matches 206; Conservative 37; Mismatches 95; Indels 3; Gaps 2;

```

```

QY 3 LSTQPDMDHMLRDGDVLAHVSHSAHFGRALVWPNLKPPTTTAAVAAYEBAIILKSL 62
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 22 LSLARPDMHMLRDGMLAAVLPHTARQGRALVWPNLKPPTTTAAQAYRERILAAV 81
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 63 PVSDPRLMTLYLTDTTSPMEIKLARESGVGVKLYPAGATTNSDGVTDLFGKCLPV 122
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 82 PAGMTPEPLMTLYLTDTTPDEIRARASGVHGVKLYPAGATTNSDGVTDL-AKCAKT 140
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 123 LOEWEHNMPLVHGEVTNPEVDMFDEKVFIEVLRPLVQKPOLKVMEHVTITDAVK 182
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 141 LEAQETGMPLVHGEVTDASIDLFDEKVFIDVMTPLRDPGLKVFEEHITTKAAD 200
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 183 FVE--SCTEGFVAATVTPQHLVLRNSLFOGGLQPHNYCLPVLKREIHRALVSAVTSGS 240
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 201 YVRDADAAPGLIGCTITTAHMLVLRNALFVGGIRPHYCLPVLRKREHRAVALVEAASGN 260
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 KRPFGLTDSAPHDRRRKESCGCAGIYNAPVALSVKAVEKENALDKLEAFISFNGPDF 300
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 261 PRFPLGTDSAPHADAKETACGAGCYTALHLELVYAEAFDTAGALDKLBGFASFQADF 320
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 YGLPRNSKITLSTKTPKVPESFSYASGDIIPMAGEMLDW 341
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 321 YGLPRSAETVTLRREPWELPREIFAGETPVVPLRGGETIGW 361
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 5
US-10-282-122A-50426
; Sequence 50426, Application US/10282122A
; Publication No. US200400291239A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseilbeck, Robert

```

```

; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Mall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50426
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Burkholderia mallei
US-10-282-122A-50426

```

```

Query Match      58.4%; Score 1058; DB 15; Length 352;
Best Local Similarity 58.4%; Pred. No. 5.6e-99;
Matches 199; Conservative 45; Mismatches 95; Indels 2; Gaps 1;

```

```

QY 3 LSTQPDMDHMLRDGDVLAHVSHSAHFGRALVWPNLKPPTTTAAVAAYEBAIILKSL 62
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 9 LTLARPDMHMLRDGMLAAVLPHTARQGRALVWPNLKPPTTTAAQAYRERILAAV 68
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 63 PVSDPRLMTLYLTDTTSPMEIKLARESGVGVKLYPAGATTNSDGVTDLFGKCLPV 122
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 69 PAGMTPEPLMTLYLTDTTPDEIRARASGVHGVKLYPAGATTNSDGVTDLFGKCAKT 128
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 123 LOEWEHNMPLVHGEVTNPEVDMFDEKVFIEVLRPLVQKPOLKVMEHVTITDAVK 182
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 129 LEAQETGMPLVHGEVTDSDIDLFDEKVFIDVMTPLRDPGLKVFEEHITTKAAD 188
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 183 FVE--SCTEGFVAATVTPQHLVLRNSLFOGGLQPHNYCLPVLKREIHRALVSAVTSGS 240
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 189 YVRDADAAPGLIGCTITTAHMLVLRNALFVGGIRPHYCLPVLRKREHRAVALVEAASGN 248
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 KRPFGLTDSAPHDRRRKESCGCAGIYNAPVALSVKAVEKENALDKLEAFISFNGPDF 300
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 249 PRFPLGTDSAPHAGAEACGAGCYTALHLELVYAEAFDTAGALDKLBGFASFQADF 308
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 YGLPRNSKITLSTKTPKVPESFSYASGDIIPMAGEMLDW 341
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 309 YGLPRSAETVTLRRETWELPREIDAGNGPVVPLRGGAIGW 349
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 6
US-10-282-122A-77674
; Sequence 77674, Application US/10282122A

```

```

Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EPIRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 77674
LENGTH: 350
TYPE: PRT
ORGANISM: Vibrio cholerae
US-10-282-122A-77674

Query Match
Best Local Similarity 57.4%; Score 1039.5; DB 15; Length 350;
Matches 200; Conservative 45; Mismatches 91; Indels 3; Gaps 3;
3 LSTTODDMMHMLRGDYLKAVVSHSAHFGRAIWNPNKRPITTTAAVAAYRAIISL 62
12 LITTPDDMMHMLRGDYLKAVVSHSAHFGRAIWNPNKRPITTTAAVAAYRAIISL 62
63 PVDSPNDMLTYLTDTPSPMEIKLARESOVFGVKLYPAGATTNSDGGVTDLFG 71
72 P-QAHFELMALYLTDNTPSEIRAKKASGVAAKLYPAGATTNSDGGVTDLFG 122
123 LOEVENHMLLVHGVTPPEVDMFDRKVFIEVLPVOKPOLKVMHEVTTIDAVK 122
130 LOAMOEVMMLLVHGVTPPEVDMFDRKVFIEVLPVOKPOLKVMHEVTTIDAVK 129
183 FVSCSEGVAATVTPQHLVLRNSLFOGGLQHPNLYCLPVKKEIHREALVSAVTSKSR 189
190 FVQAGADN-VAATTTAHHILFNRNMLVGGIRHFFCLPLKRAVCHQALVVAATSSSK 242
243 FLGLTDSAPHDRRREKSCGAGIYUAPVALSYAVKFEKENALDKLBAFTSFRGDPFG 248
249 FLGLTDSAPHAKGRKKAACGAGIYUAPVALSYAVKFEKENALDKLBAFTSFRGDPFG 302
303 LPRNNSKIKLSTKTPWKYPRSEFVSAGDIIPMFAGEMLDW 341

```

```

Db
309 LPRNDETTLTKQAMPVASEMPSGSDIVPIRAGENIEW 347

RESULT 7
US-10-282-122A-51386
Sequence 51386, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EPIRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 51386
LENGTH: 354
TYPE: PRT
ORGANISM: Bordetella pertussis
US-10-282-122A-51386

Query Match
Best Local Similarity 55.7%; Score 1008.5; DB 15; Length 354;
Matches 196; Conservative 46; Mismatches 97; Indels 5; Gaps 3;
2 ELSTODDMMHMLRGDYLKAVVSHSAHFGRAIWNPNKRPITTTAAVAAYRAIISL 58
8 ELITRPDDMMHMLRGDYLKAVVSHSAHFGRAIWNPNKRPITTTAAVAAYRAIISL 58
59 LKSLPVD-SDFNPMLTYLTDTPSPMEIKLARESOVFGVKLYPAGATTNSDGGVTDLFG 117
68 LRAAGDGTAAFTPLMTLYLTDNTPSEIRAKKASGVAAKLYPAGATTNSDGGVTDLFG 127
118 KCLPVLOEVENHMLLVHGVTPPEVDMFDRKVFIEVLPVOKPOLKVMHEVTTIDAVK 127
128 KCGAALLALECGMPLLVHGVTPPEVDMFDRKVFIEVLPVOKPOLKVMHEVTTIDAVK 177
178 IDAVKFEVSCSEGVAATVTPQHLVLRNSLFOGGLQHPNLYCLPVKKEIHREALVSAVTSKSR 237
188 RBGAHYVD-ABGTAATITTPQHLVLRNSLFOGGLQHPNLYCLPVKKEIHREALVSAVTSKSR 246

```

Qy	238	SGSKRPETGSDA	PHRRRREKSCCGAGI	YNA	PVAL	SLYAVAFKEKNAL	LDLGEFTSG	297
Db	247	SGSRFFFLGDS	PHRRGLKEHKGCGACG	TYLHM	MELYATRA	FDVAGLDP	LDGEFAFHG	306
Qy	298	PDFYGLPRNNSK	IKLSTPKVPESEFV	SGDIIM	PMFAGELMD			341
Db	307	PDFYGLPRNNG	LTLTRREAV	ELPAEVA	FDDTLVL	PSGSEISGM		350

RESULT 8
US-10-282-122A-65811

```

1  /
2  / TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
3  /
4  / APPLICANT: ELITRA.034A
5  /
6  / FILE REFERENCE: ELITRA.034A
7  /
8  / CURRENT APPLICATION NUMBER: US/10/282,122A
9  /
10 / CURRENT FILING DATE: 2003-02-20
11 /
12 / PRIOR APPLICATION NUMBER: 60/191,078
13 /
14 / PRIOR FILING DATE: 2000-03-21
15 /
16 / PRIOR APPLICATION NUMBER: 60/206,848
17 /
18 / PRIOR FILING DATE: 2000-05-23
19 /
20 / PRIOR APPLICATION NUMBER: 60/207,727
21 /
22 / PRIOR FILING DATE: 2000-05-26
23 /
24 / PRIOR APPLICATION NUMBER: 60/230,335
25 /
26 / PRIOR FILING DATE: 2000-09-06
27 /
28 / PRIOR APPLICATION NUMBER: 60/230,347
29 /
30 / PRIOR FILING DATE: 2000-09-09
31 /
32 / PRIOR APPLICATION NUMBER: 60/242,578
33 /
34 / PRIOR FILING DATE: 2000-10-23
35 /
36 / PRIOR APPLICATION NUMBER: 60/253,625
37 /
38 / PRIOR FILING DATE: 2000-11-27
39 /
40 / PRIOR APPLICATION NUMBER: 60/257,931
41 /
42 / PRIOR FILING DATE: 2000-12-22
43 /
44 / PRIOR APPLICATION NUMBER: 60/267,636
45 /
46 / PRIOR FILING DATE: 2001-02-09
47 /
48 / PRIOR APPLICATION NUMBER: 60/269,308
49 /
50 / PRIOR FILING DATE: 2001-02-16
51 /
52 / Remaining Prior Application data removed - See File Wrapper or PALM.
53 /
54 / NUMBER OF SEQ ID NOS: 78614
55 /
56 / SOFTWARE: PatentIn version 3.1
57 /
58 / SEQ ID NO 65811
59 /
60 / LENGTH: 344
61 /
62 / TYPE: PRT
63 /
64 / ORGANISM: Neisseria meningitidis
65 /
66 / US-10-282-122A-65811

```

	Query Match	54.2%	Score 983;	DB 15;	Length 344;
	Best Local Similarity	56.3%;	Pred. No. 2.4e-91;		
	Matches 192;	Conservative 55;	Mismatches 88;	Indels 6;	Gaps 5
Qy	3 LSI TD DDHMLHRDGDVLKAVVSHAHFGRALVMNPKIPITTTAAVAAYREALIKSL	62			
Dd	4 LTIRPD DHHLRDRDDALKAPARYARQMGAIVIMPNKKPVVSADALAKYRIMATL	63			
Qy	63 PVDSDNPPLMTLYLTPTTS PMELKLARESGYVGVLTPAGATPSSQDSGVTLFGGCLPY	122			
Dd	64 PGCSAEPRLTLLTLTGQAIFPALREKAAGIVAFGLYPAGATTNSDSGVTLDF-KLIPIV	121			
Qy	123 IQEMVEHNMPLVHGGEVTINPEVDMPDKRKVIETTVLRPLVKRFPOLKYMMEHTTIDAVK	182			

Db	122	LEEMANQGLFLVLHGVEVTDPEIDTIDREAAFTGRVMKRVLAQVENVLKVEVNHITTAEAR	181
Oy	183	FVESCTEGEVAAVTTPQHLVLNRNSLFGQGIOPHYCLPVLKREIHREBALVAVTSCK-	241
		182 LVLEAGDN-VAAITVTPQHLNRNLLVGVGRPHHFCPLVKRETHQALVAAVT-CEKA	239
Db	242	-REFLGTDSAPHDRRRKECCGCGAGIYAAPVALASLYAAKVEKENALDKLEAFSTFENPDF	300
Oy	240	HKFFLGTDSSAPHAASKAKENACGACAMFSAMTIEIYAAVEFEKAGLDLKEAFSAKNGARF	299
Db	301	YGLPRNNSKIULSKTTPMKVPPESFVASGDIIIPMAGELTMD	341
Oy	300	YGIPEVNDITTLVKOSQTVAPISIPGDEGLVPMRGGEGIGW	340

RESULT 9
US-10-282-122A-65238
; Sequence 65238, Application US/10282122A
; Publication No. US20040029129A1

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

; ORGANISM: Neisseria gonorrhoeae
US-10-282-122A-65238

Query Match          54.0%; Score 979; DB 15; Length 344;
Best Local Similarity 56.3%; Pred. No. 6, 2e-91;
Matches 192; Conservative 55; Mismatches 88; Indels 6; Gaps 5

QY      3 LSTLPDDMHLLRGDYLKAVNSAHNFGAIIVPMLKPITTTAAAVVREALIKSL 62
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      4 LTIRRDDHNLRLRGDALKAAPATPAOMGAIVMPNPKRPVSVADALAKATIMALL 63

```


QY 63 PVDSDFNPLMTLYLTDTTSPMEIKLARESOVFGVKLYPAGATTNSODGVTDLFGKCLPV 122
DB 64 PGCSAFEPPLMTLYLTDOATPELVREAKAGIV-APKLYPAGATTNSOSGVTDLF-KLIPV 121
QY 123 LOEMENHNPMLLVHGEVTNPEVDMFDRKVFIEYTLRPLVOKFPOLKVMEHTTTIDAVK 182
DB 122 LEMMEKQIGPLVHGEVTDPEIDIFDREAAFIKGVMPVLAQVPMKLVPEHITTTIDAVK 181
QY 183 FVESCIEGFVAATVPQHLVLRNSLFOGGLQPHNYCLPVLRKREIHEALVSATVSGSKR 241
DB 182 LVLEAGDN-VAATVPQHLINRNDLVGGVRPHHFLPVLRKREIHEALVSATVSGSKR-GEKA 239
QY 242 -RFLGTDASAPHRRRKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGDPDF 300
DB 240 HKFGLGDSAPHAASAKENACGCGAGMFSAMTALIELVAEVEKAGALDKLEAFSAKNGARF 299
QY 301 YGLPRNNSKI.KLSTPKVPESFSYASGDIIPMFAGEMLDW 341
DB 300 YGIPENADTITIVKQSGTVPASVPYGDGELVPMRAGGETGW 340

RESULT 10

US-10-282-122A-59728
Sequence 59728, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Travick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282.122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 59728
LENGTH: 348
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59728

Query Match 54.0%; Score 978; DB 15; Length 348;
Best Local Similarity 55.8%; Pred. No. 8e-91;

Matches 189; Conservative 46; Mismatches 102; Indels 2; Gaps 2;
QY 3 LSTIQPDWHLHRLDGDVLKAVVSHSAHFGRAIWNPNKRPITTTAAVAAREAILKSL 62
DB 8 LKIRRPDMHILHRDDMLKTVTPYTSFYGALVWPNLPVPTVVAALAAVROKIMAV 67
QY 63 PVDSDFNPLMTLYLTDTTSPMEIKLARESOVFGVKLYPAGATTNSODGVTDLFGKCLPV 122
DB 68 PACHDFPLMTCYLTSLSLPAELERGFNEGFTAKLYPANATTNSHGVTST-DALMPV 126
QY 123 LOEMENHNPMLLVHGEVTNPEVDMFDRKVFIEYTLRPLVOKFPOLKVMEHTTTIDAVK 182
DB 127 LEMMEKQIGPLVHGEVTDPEIDIFDREARFIETMEPLRQRLPELKVFPHITTKAAAE 186
QY 183 FVESCIEGFVAATVPQHLVLRNSLFOGGLQPHNYCLPVLRKREIHEALVSATVSGSKR 242
DB 187 YVADGNE-LLAATITPQHLMFNRNMLVGGIRPHLYCLPVLRNITHOQALRELVASGFSR 245
QY 243 FFLGTDASAPHRRRKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFSTFNGDPDF 302
DB 246 AFGTDASAPHAARREASGCGAGCFNAPTALGSTATVFEENMALQHFARCSINGPRFYG 305
QY 303 LPRNNSKI.KLSTPKVPESFSYASGDIIPMFAGEMLDW 341
DB 306 LRVNESYVELVREETTVDISTALPNDLVLPFLAGETVRW 344

RESULT 11

US-10-282-122A-66489
Sequence 66489, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Travick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282.122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 66489
LENGTH: 348


```
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 67595
/ LENGTH: 348
/ TYPE: PRF
/ ORGANISM: Pseudomonas putida
US-10-282-122A-67595
```

```
Query Match 53.1%; Score 962; DB 15; Length 348;
Best Local Similarity 53.7%; Pred. No. 3,5e-89;
Matches 182; Conservative 46; Mismatches 109; Indels 2; Gaps 2;
```

```
QY 3 LSIITPPDDMHLHLDGVDLKAIVSHSAHFGRALVMPNLPKPIITTTAAVAAYREALIKSL 62
DB 5 LTLRLPDDMHLHLDGAVLPHTVGAVARTFARALIMPLVPPVNAVTEAGYRERILAAAR 64
QY 63 PVDSDPNPLMTLYLTDTSPEIKLARESQQVFGVKLYPAGATTNSODGVTDLFCKCLPV 122
DB 65 PAGRFPPLMTLYLTDRTSPEDVRAAKASGFVYAAKLYPAGATTNSDGVTSI-DNIFPA 123
QY 123 LQEWENHMPPLVHGEVTPNPEVDMFDEKVFIEFTVLRPLVOKFPOLKVMHEHTTIDAVK 182
DB 124 IEALAEVGMPLLVHGEVTRSEIDVDRKRFIDEMRVRVERFPTLVVEHITTSDAQ 183
QY 183 FVESTCEFVAATYTPQHLVLRNSLFOGGLQPHNYCLPVLRKRIHREALVSAVTSQSKR 242
DB 184 FVTEAPAN-VGATITTAHLVLRNMLVGGIRPHFYCLPIIKRTHQVALLDAATSNPK 242
QY 243 FFLGDSAPHDRRREKSCGAGIYNAPVALSVYAKFEKENALDKLEATSNRGPDPFG 302
DB 243 FFLGDSAPHDRRREKSCGAGIYNAPVALSVYAKFEKENALDKLEATSNRGPDPFG 302
QY 303 LPRNNSKIKSKTPWKVPESFVSAGDIIPMFAGEMLDW 341
DB 303 LPRNNSKIKSKTPWKVPESFVSAGDIIPMFAGEMLDW 341
DB 303 LPRNNSKIKSKTPWKVPESFVSAGDIIPMFAGEMLDW 341
```

RESULT 14

```
US-10-282-122A-69927
/ Sequence 69927, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
```

```
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 69927
/ LENGTH: 347
/ TYPE: PRF
/ ORGANISM: Pseudomonas syringae
US-10-282-122A-69927
```

```
Query Match 52.9%; Score 959; DB 15; Length 347;
Best Local Similarity 55.2%; Pred. No. 7e-89;
Matches 187; Conservative 44; Mismatches 106; Indels 2; Gaps 2;
```

```
QY 3 LSIITPPDDMHLHLDGVDLKAIVSHSAHFGRALVMPNLPKPIITTTAAVAAYREALIKSL 62
DB 5 LTLRLPDDMHLHLDGAVLPHTVGAVARTFARALIMPLVPPVNAVTEAGYRERILAAAR 64
QY 63 PVDSDPNPLMTLYLTDTSPEIKLARESQQVFGVKLYPAGATTNSODGVTDLFCKCLPV 122
DB 65 PAGRFPPLMTLYLTDRTSPEDVRAAKASGFVYAAKLYPAGATTNSDGVTSI-DKIFPA 123
QY 123 LQEWENHMPPLVHGEVTPNPEVDMFDEKVFIEFTVLRPLVOKFPOLKVMHEHTTIDAVK 182
DB 124 IEALAEVGMPLLVHGEVTRSEIDVDRKRFIDEMRVRVERFPTLVVEHITTSDAQ 183
QY 183 FVESTCEFVAATYTPQHLVLRNSLFOGGLQPHNYCLPVLRKRIHREALVSAVTSQSKR 242
DB 184 FVTEAPAN-VGATITTAHLVLRNMLVGGIRPHFYCLPIIKRTHQVALLDAATSNPK 242
QY 243 FFLGDSAPHDRRREKSCGAGIYNAPVALSVYAKFEKENALDKLEATSNRGPDPFG 302
DB 243 FFLGDSAPHDRRREKSCGAGIYNAPVALSVYAKFEKENALDKLEATSNRGPDPFG 302
QY 303 LPRNNSKIKSKTPWKVPESFVSAGDIIPMFAGEMLDW 341
DB 303 LPRNNSKIKSKTPWKVPESFVSAGDIIPMFAGEMLDW 341
DB 303 LPRNNSKIKSKTPWKVPESFVSAGDIIPMFAGEMLDW 341
```

RESULT 15

```
US-10-282-122A-78404
/ Sequence 78404, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using 6w model

Run on: February 12, 2005, 06:47:55 ; Search time 29 Seconds
(without alignments)
890.641 Million cell updates/sec

Title: US-10-070-277-2

Perfect score: 1812
Sequence: 1 MELSTQDDMHLDGDPV.....SGDITPMFAGMLDWIPAPLV 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgm2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgm2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgm2_6/prodata/1/iaa/6A_COMB.pep:*
4: /cgm2_6/prodata/1/iaa/6B_COMB.pep:*
5: /cgm2_6/prodata/1/iaa/PCTUS_COMB.pep:*
6: /cgm2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	978	54.0	352	4	US-09-489-039A-7589 Sequence 7589, App
2	975	53.8	398	4	US-09-252-991A-17491 Sequence 17491, A
3	952.5	52.6	346	4	US-09-328-352-5541 Sequence 5541, App
4	946	52.2	355	4	US-09-540-236-5330 Sequence 2330, App
5	930	51.3	356	4	US-09-543-681A-4285 Sequence 4285, App
6	522.5	28.8	365	4	US-09-248-796A-18106 Sequence 18106, A
7	132	7.3	734	4	US-09-902-540-11399 Sequence 11399, A
8	119.5	6.6	430	4	US-09-107-532A-5107 Sequence 5107, App
9	116.5	6.4	448	4	US-09-134-000C-6087 Sequence 6087, App
10	103.5	5.7	422	4	US-09-583-110-3844 Sequence 3844, App
11	103.5	5.7	426	4	US-09-107-433-3349 Sequence 3349, App
12	102.5	5.7	371	4	US-09-710-279-628 Sequence 628, App
13	102.5	5.7	425	4	US-09-710-279-2604 Sequence 2604, App
14	102.5	5.7	427	3	US-09-134-001C-4803 Sequence 4803, App
15	97	5.4	244	3	US-08-469-318-126 Sequence 126, App
16	97	5.4	244	3	US-08-468-609A-126 Sequence 126, App
17	97	5.4	244	3	US-08-446-872A-126 Sequence 126, App
18	97	5.4	244	4	US-08-762-227A-126 Sequence 126, App
19	97	5.4	244	5	PCT-US95-01185-126 Sequence 126, App
20	96.5	5.3	699	4	US-09-438-185A-506 Sequence 506, App
21	94	5.2	244	3	US-08-469-318-125 Sequence 125, App
22	94	5.2	244	3	US-08-468-609A-125 Sequence 125, App
23	94	5.2	244	3	US-08-446-872A-125 Sequence 125, App
24	94	5.2	244	4	US-08-762-227A-125 Sequence 125, App
25	94	5.2	244	5	PCT-US95-01185-125 Sequence 125, App
26	94	5.2	285	3	US-08-875-533-69 Sequence 69, App
27	89.5	4.9	681	4	US-09-252-991A-18889 Sequence 18889, A

28	89.5	4.9	734	4	US-09-809-665A-117 Sequence 117, App
29	89	4.9	244	3	US-08-469-318-127 Sequence 127, App
30	89	4.9	244	3	US-08-468-609A-127 Sequence 127, App
31	89	4.9	244	3	US-08-446-872A-127 Sequence 127, App
32	89	4.9	244	4	US-08-762-227A-127 Sequence 127, App
33	89	4.9	244	5	PCT-US95-01185-127 Sequence 127, App
34	89	4.9	285	3	US-08-875-533-70 Sequence 70, App
35	89	4.9	458	4	US-09-497-585A-8 Sequence 8, App
36	89	4.9	458	4	US-09-916-501A-14 Sequence 14, App
37	89	4.9	469	4	US-09-540-236-2593 Sequence 2593, App
38	88.5	4.9	627	4	US-09-360-545-32 Sequence 32, App
39	88.5	4.9	868	4	US-09-949-016-11723 Sequence 11723, A
40	88	4.9	191	2	US-08-438-439C-22 Sequence 22, App
41	88	4.9	244	3	US-08-469-318-136 Sequence 136, App
42	88	4.9	244	3	US-08-468-609A-136 Sequence 136, App
43	88	4.9	244	3	US-08-446-872A-136 Sequence 136, App
44	88	4.9	244	4	US-08-762-227A-136 Sequence 136, App
45	88	4.9	244	5	PCT-US95-01185-136 Sequence 136, App

ALIGNMENTS

```
RESULT 1
US-09-489-039A-7589
Sequence 7589, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489, 039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117, 747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 7589
LENGTH: 352
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7589

Query Match      54.0% Score 978; DB 4; Length 352;
Best Local Similarity 55.8% Pred. No. 1.8e-98;
Matches 189; Conservative 46; Mismatches 102; Indels 2; Gaps 2;

QY 3 LSTQDDMHLDGDPVLSHSHHRAIVNPNLKPITTTAAVAVEALDKSL 62
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 12 LKTRPDDMHLDGDPVLSHSHHRAIVNPNLKPITTTAAVAVEALDKSL 71
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 63 PVDSPNPLMTLYLTDTTSFMEIKLAESQVFGVKLYPAGATTNSQDVTDLFGKCLPV 122
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 72 PACHDFPLMTCTLTDSLDPAELERGFNEGVTFAAKLYPANATTNSHGVTST-DAIMPV 130
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 123 LQEMVHNMPLVHGEVTNEVVMFDEKPIETVLRLVOKFPQLVWNEHVTITDAVK 182
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 131 LEMNEKIMPLVHGEVTHAEIDIFDEKAPFETVMEPKLORGLGVVEHITTKDAE 190
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 183 FVSSCTGEFAATVTPQHLVLRNSLFOGLQPHNYCLPYLKEIRHREALVSATSGSKR 242
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 191 YVRDGNL-LIAAITTPHLMFNRRNMLVGGIRHLVCLPVLKKNIHQALRELVASGFSR 249
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 243 FLITGDSAPHRRRRKEKSCGACIYNAPVALSVYAKVFEKENALDKLEAFTSGFDPFYG 302
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 250 AFLGTGDSAPHRRRRKEKSCGACIYNAPVALSVYAKVFEKENALDKLEAFTSGFDPFYG 309
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 303 LPRNNSKIKSKTPKWKYPSBSFYASGDIIIMFAGEMLDW 341
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 310 LPVNESYVELVRETTVDVSDIALPNDTLVPLAGETVRW 348
    |||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Query Match	52.6%;	Score 952.5;	DB 4;	Length 346;
Best Local Similarity	53.8%;	Pred. No. 1.1e-95;		
Matches 183;	Conservative 57;	Mismatches 97;	Indels 3;	Gaps 3;
QY	3	LSITQPDWHLRLRDGVIAKAVSHSAHFGRALVNMENLKPITTTAAAVAREALISL	62	

[illegible]

RESULT 5
US-09-543-681A-4285


```
; Sequence 4285, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4285
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4285

Query Match      51.3%; Score 930; DB 4; Length 356;
Best Local Similarity 52.8%; Pred. No. 3,4e-93;
Matches 179; Conservative 51; Mismatches 107; Indels 2; Gaps 2;

QY      3  LSTGDDMHLLHRRDGDVLAAYVSHSAHFRGALVMPNKLKRPITTTAAAVAYREALIKSL 62
      15  LTRRPDDMHVHRRDDMDLMTVPTYSRYGRALVMPNLPITTTTAAARVDRIRKAAI 74
      63  PVSDPNPLMTLYLTDITTSFMEIKLARESQVFGVKLYPAGATTNSQDVTDFGKCLPV 122
      75  PSGDKFEPLMTCTLTSTSEVGEQFLQGVFTACKLYPANATNSHGVSDI-NKIYPI 133
QY      123  LQEMVHNHMLVHGEVTNEVDMFDEKVFIEVLRLVQKRPOLKVMNEHTTIDAVK 182
      134  LSYMERKGMPLIHGEVTASDIDIPEAREFIDNVAPVKOPALKIVEHEHTTTEMAQ 193
QY      183  FVSSCEGFAATVTPQHLVLNNSLQGLQPHNYCLPVLKEIHREALVSAVTSQSKR 242
      194  YVLEGNE-FLGATITTPQHLNFRNRMHVLGVKPHLYCLPLKRNVOEALROVAASHSR 252
QY      243  FLIGTDSAPHRRRKCCSCGAGIYNAPVALSVYAKVEKENALDKLEAFTSFNGDPFYG 302
      253  FLIGTDSAPHLQRRKSSCCGAGVFNAPFTLAAVASFKMLNLSHEAFCSLNGPRFYG 312
QY      303  LPRNNSKIKLSKTPWKVPESFSYASGDIIIPFAGEMLDW 341
      313  LPVNEGTTITLTKSVTAPAEIMSGDEALIFLANEDIHW 351
DB

RESULT 6
US-09-248-796A-18106
; Sequence 18106, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196,132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18106
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-18106

Query Match      28.8%; Score 522.5; DB 4; Length 365;
Best Local Similarity 38.3%; Pred. No. 1.6e-48;
Matches 139; Conservative 57; Mismatches 132; Indels 35; Gaps 15;
```

```
QY      1  MELSTGQDDMHLLHRRDGDVLAAYVSHSAHFRGALVMPNKLKRPITTTAAAVAYREALI 58
      11  IELGITA--DMVHLRDSMCE-LITPTRTGGIALSYMPNLPVPIITTKQOVVNHKTL 67
QY      59  LKSLPVSDPNPLMTLYLTDITTSFMEIKLARE-SQVFGVKLYPAGATTNSQDVT-DLIF 116
      68  SKLAP-QTTF--LMSFYLSKDLTP---ELIEECADLIHGVCKYCPAGVTTNSKGVDPNDF 121
QY      117  GKCLPVQLQEMVHNHMLVHGE---VMPNVDMPREKVFIETVRLVQKRPOLKVM 172
      122  SSFYPLFEMQKRGVLNTHGKPGTSEDEINVAEKFV--PALKULHADFPRLKIL 180
QY      173  EHVTITDAVKFVESCT-----EGFVAATVTPQHLVLNNSLQGLQPHNYCLPVLK 225
      181  ECHTSDAVEYTRBELNSNKPQGDLEFVGATITTAHHLYLIIDN-WAG--NPINFCVKVAF 237
QY      226  EIHREALVSAVTSQSKRFFLGTDSDAPHRRRKCCSCG-CAGIYNAPVALSVAKVEKEN 284
      238  QDKRKALVDAAATSGEPMFFGSDSADPHIAKQTHGVGACAGVYTOSHALAYVAEIPDKAN 297
QY      285  ALDKLEAFTSFNGDPFYG-----PRNNSKIKLSKTPWKVPESFSYASGDIIIPFAGEM 338
      298  KLDNLKKFVSDNGIKFYGLSDILKKHKNESVWLVRDNQVPEIIGNDVEVVPFPAGET 357
QY      339  LDW 341
      358  LKY 360
DB

RESULT 7
US-09-902-540-11399
; Sequence 11399, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldmann, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequence and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 11399
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-11399

Query Match      7.3%; Score 132; DB 4; Length 734;
Best Local Similarity 22.5%; Pred. No. 2.9e-05;
Matches 81; Conservative 46; Mismatches 115; Indels 118; Gaps 18;

QY      10  DMHLLHLD-GD-----VLKAVSHSAHFRGALVMPNKLKRPITTTAAAVAYREALISLP 63
      359  DLHVLHLEPGEBEKGYLTGCRAAVAGGTAIVAMVPYTK----- 397
QY      64  VDSDFNPLMTLYLTDITTSFMEIKLARESQVFGVKLYPAGATTNSQDGT----- 111
      398  -----VNNDNAMTVELVLR-ARAAGLCHVYPAGAITGLGABELAEWGEIVSA 445
QY      112  ---VTDLFGKCLPV-----LQEMVHNHMLVH----- 136
      446  GCVAITD-DGR--FVNNASLMRRTLYQATQFDVPMVHEEDLTLASAGANHEGTTSTRLG 502
QY      137  --GEVTMPYVDMPREKVFIETVRLVQKRPOLKVMNEHTTIDAVKVF-ESCTGEF-V 192
      503  LRGIIPAFAEYAMVARDVLEET-----KGRLVAAVSCGSRRLIEAKRSGLRV 553
QY      193  AATVTPQHLVLNNSLQGLQPHNYCLPVLKREIHREALVSAVTSQSKREFLGTDSPVH 252
```

Db 554 TCVEAPHHFILDDRAB--GDYDTHAKMAPPLADTDVQALREAVDGTVD-AIATDAPH 610
QY 253 D--RRREKSCGAGIYNAPVALS-----VYAKVEKENALDKLEAFSTFNGP-DYGLP 304
Db 611 GVDKQVEKFGINGIVLEIETALGLTELHVAGVLTARAVELLT-----HGPAKAGLP 665

RESULT 8
US-09-107-532A-5107
Sequence 5107, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 60/095,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5107:
SEQUENCE CHARACTERISTICS:
LENGTH: 430 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc.feature
LOCATION: (B) LOCATION 1...430
SEQUENCE DESCRIPTION: SEQ ID NO: 5107:
US-09-107-532A-5107

Query Match
Best Local Similarity 23.3%; Score 119.5; DB 4; Length 430;
Matches 67; Conservative 43; Mismatches 114; Indels 63; Gaps 14;

QY 10 DWHLHRLD-GDYLKAVV-----SHSAHFGRATVWPNLKP-PITTTAAVAV-----REAI 58
Db 62 DVVHFRFBPGFTYKETIGAGTSARGGFTTVCAMPNLPVDTAEKLRQVYDIIKDAV 121
QY 59 LKSLPVDSDFNPLMTLYLTDITSPMEIKLARSQVVFVKLYPAGATTNSQDGV-TDLFG 117
Db 122 VKVLQY-----APTENTL-RSEKLVQDEALIERGAFAPFINDGVQTAG 164
QY 118 KCLPVLOEVENHNPVLVHGEVTNPEVDMF-----DREKVFIEVLRP-----160
Db 165 TWYLAKEAAKNNKALVANH--TEDESLFGGVNHAKKAELGLPGILSTVESSQIARD 221

QY 161 -LVQKFPQKVMMEHTTIDAVKVESCTEG--FVAATVTPQHLVLRNSLFGGLQPHN 217
Db 222 LLLEAGGVNHVCHVSTKESVRVIRDAKKAIGHVAEVSPPHLLIDEDI-----PBD 275
QY 218 YCL-----PVLKREIHRBALVSAVTSKRFPLGTDSAPHRKXCS 260
Db 276 FGFWKKNPPLRGREDREALIEGLDGTID-CIATDHAHPGLEKXCS 321

RESULT 9
US-09-134-000C-6087
Sequence 6087, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
PRIOR FILING DATE: 1998-08-13
CURRENT APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: Patent in version 3.1
SEQ ID NO: 6087
LENGTH: 448
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-6087

Query Match
Best Local Similarity 6.4%; Score 116.5; DB 4; Length 448;
Matches 80; Conservative 39; Mismatches 124; Indels 87; Gaps 18;

QY 10 DWHLHRLD-GDYLKAVV-----SHSAHFGRATVWPNLKP-PITTTAAVAV-----REAI 58
Db 79 DVVHFRFBPGFTYKETIGSKAARAGGFTTVCAMPNLPVDTAEKLSVYDIIQDAV 138
QY 59 LKSLPVDSDFNPLMTLYLTDITSPMEIKLARSQVVFVKLYPAGATTNSQDGV-TDLFG 117
Db 139 VKVLQYAPITEELRSFVLINOKA-----LKEAR-----AFAFN-DGVGVQTAG 181
QY 118 KCLPVLOEVENHNPVLVHGEVTNPEVDMFDEKVFIEVLRPVQK-----FPOKVMME- 173
Db 182 TWYLAKEAAKNNKALVANH--TEDESLFGGVNHAKKAELGLPGILSTVEA 232
QY 174 -----HTTIDAVKVESCTEG--FVAATVTPQHLVLRNSLFG 210
Db 233 SQIARDITLAKETGVHNVCHVSTESVRVIRDAKKAIGHVAEVSPPHLLIDEDI 292
QY 211 GGLQPHNYCLPVLRKEIHRBALVSAVTSKRFPLGTDSAPHRKXCSGCGAGIYNAP 270
Db 293 BEFWKON--PPLGLADRLQALIDGLDGTID-CIATDHAHPGLEKXCS-----FLANP 343
QY 271 VAL-----SVYAKVEKENALD 287
Db 344 FGIVSETAFQILYTNFVETGIFTLQVID 373

RESULT 10
US-09-583-110-3844
Sequence 3844, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
FILE REFERENCE: PATNO-07A
CURRENT APPLICATION NUMBER: US/09/583,110
PRIOR FILING DATE: 2000-05-26
CURRENT APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30

PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 3844
LENGTH: 422
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-583-110-3844

Query Match 5.7%; Score 103.5; DB 4; Length 422;
Best Local Similarity 20.8%; Pred. No. 0.017;
Matches 71; Conservative 54; Mismatches 131; Indels 85; Gaps 16;

10 DWHLH:RD-----GDVKAIVSHSAHFGRAIVMPLKPIITTTAAVAAYREALIKS-- 61
57 DIHVHREBPQTHKEDIHTGALAAAAGFTTVMMANTSPTISDVETLQAVLSAAKEKI 116
62 -----LPVSDENPLMTLYITDTSPEIKLARESOVFGVKLYPAGATTNSODGVTDLF 116
117 NVKTVATITGNFN---GKNLTDFKALLE-----AGAVGFSDDGIPLES 156
DB 117 GKCLPVLQGVENHNMPLVHGEVTNPVDMFDEK-----VFETVLRP----- 160
DB 157 SK---LVKEMAEAKKL-----NTFISLHEEDPGLNGVLGFENENIAREHFIHICGATGV 206
QY 161 -----LVQKPOLKVMHEVTTIDAVKEVSECTEGF---VAATVTPHVLNIRNS 207
DB 207 AEYAMMARDMIAVATKAHIIQHSKESEVKVE-FAQGLGAEVTEVAAPQHSKTEAL 265
QY 208 LP-OGGLQPNHYCLPVKREIHRBALVSATSGSKREFLCTDSAPH--DRRKECGCGC 263
DB 266 LITQGSNAKKN---PPLRLSDRRAVIEGLKSGVIT-VIATDHAHHVDEKVEDITKAP 321
QY 264 AGIYNAFVALSV--YAKVFEKENALDKLAFSFGNDPY 301
DB 322 SGMTGLTSLSLGLTYLVEAGELSLMELBKMT-YNPAKLY 361

RESULT 11
US-09-107-433-3349
Sequence 3349, Application US/09107433
Patent No. 6800744
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,433
FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 3349:
SEQUENCE CHARACTERISTICS:
LENGTH: 426 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...426
SEQUENCE DESCRIPTION: SEQ ID NO: 3349:
US-09-107-433-3349

Query Match 5.7%; Score 103.5; DB 4; Length 426;
Best Local Similarity 20.8%; Pred. No. 0.017;
Matches 71; Conservative 54; Mismatches 131; Indels 85; Gaps 16;

10 DWHLH:RD-----GDVKAIVSHSAHFGRAIVMPLKPIITTTAAVAAYREALIKS-- 61
61 DIHVHREBPQTHKEDIHTGALAAAAGFTTVMMANTSPTISDVETLQAVLSAAKEKI 120
62 -----LPVSDENPLMTLYITDTSPEIKLARESOVFGVKLYPAGATTNSODGVTDLF 116
121 NVKTVATITGNFN---GKNLTDFKALLE-----AGAVGFSDDGIPLES 160
DB 121 GKCLPVLQGVENHNMPLVHGEVTNPVDMFDEK-----VFETVLRP----- 160
QY 117 GKCLPVLQGVENHNMPLVHGEVTNPVDMFDEK-----VFETVLRP----- 160
DB 161 SK---LVKEMAEAKKL-----NTFISLHEEDPGLNGVLGFENENIAREHFIHICGATGV 210
QY 161 -----LVQKPOLKVMHEVTTIDAVKEVSECTEGF---VAATVTPHVLNIRNS 207
DB 211 AEYAMMARDMIAVATKAHIIQHSKESEVKVE-FAQGLGAEVTEVAAPQHSKTEAL 269
QY 208 LP-OGGLQPNHYCLPVKREIHRBALVSATSGSKREFLCTDSAPH--DRRKECGCGC 263
DB 270 LITQGSNAKKN---PPLRLSDRRAVIEGLKSGVIT-VIATDHAHHVDEKVEDITKAP 325
QY 264 AGIYNAFVALSV--YAKVFEKENALDKLAFSFGNDPY 301
DB 326 SGMTGLTSLSLGLTYLVEAGELSLMELBKMT-YNPAKLY 365

RESULT 12
US-09-710-279-628
Sequence 628, Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: P03480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 628
LENGTH: 371
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: amino acid sequence
US-09-710-279-628

Query Match 5.7%; Score 102.5; DB 4; Length 371;
Best Local Similarity 22.7%; Pred. No. 0.018;

```

RESULT 13
US-09-710-279-2604
; Sequence 2604, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS.
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2604
; LENGTH: 425
; TYPE: PRT
; ORGANSIM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; US-09-710-279-2604

```

```

/ Sequence 4803, Application US/09134001C
/ Patent No. 6380370
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
/ TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: CPG-007
/ CURRENT APPLICATION NUMBER: US/09/134,001C
/ PRIOR APPLICATION NUMBER: 1998-08-13
/ PRIOR FILING DATE: 1997-11-08
/ PRIOR APPLICATION NUMBER: US 60/064,964
/ PRIOR FILING DATE: 1997-08-14
/ NUMBER OF SEQ ID NOS: 5674
/ SEQ ID NO 4803
/ LENGTH: 427
/ TYPE: PRT
/ ORGANISM: Staphylococcus epidermidis
/ US-09-134-001C-4803

```

```

Query Match      5.7%; Score 102.5; DB 3; Length 427;
Best Local Similarity 22.7%; Pred. No. 0.022;
Matches 65; Conservative 40; Mismatches 114; Indels 67; Gaps 15

QY      10 DWHLRD--GDVLKAVS--SSAHNFGRLV--MPLNKPRITTTAAVAAREALSLP 63
DB      59 DVVHILREPGEHKETTESGTAAARGFTTVCMPNRPVPDIVEHRELRLRISTEAQ 118
QY      64 VDSDFNPMLTLVLTDTSTPMEL---KLARESQVFVGKLYPAGATNSODGV-TDLFGK 118
DB      119 V----RVLPYAATIKRKQGTELVDEKULREVFNF-----TDDGCVQTASM 162
QY      119 CLFVLQENVEHNMPLLIVHGEVNPEVDMPDEKYF---IETVLRIWQKPPQLKVNE-- 173
DB      163 MPAAMKAQAKVKKPIVAFCE-----DNSLIYGAMHGKRSSEELGIPIPTIAESV 213
QY      174 -----HTTIDAKVFVESCTEG--FAAATVPOLHLINRSLFOGC 212
DB      214 QIARDVLAEATGCHYHVCHGSTKESRAVRIRDAKKAGIHVAEETPHHLLLTENDV-PG- 271
QY      213 LDPNNYCL-PVLKEIHREALVSATSGSKRRFLGTDAPDRRK 257
DB      272 -DSDNYKNMPPIRSNEDREALLEGLDGITD-CIADTHAPHAKEEK 315

RESULT 15
US-08-469-318-126
; Sequence 126, Application US/08469318
; Patent No. 6022535
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis Fusion
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 196
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,318
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/446,872
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 126:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULAR TYPE: protein

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: February 12, 2005, 05:12:50 ; Search time 92 Seconds
(without alignments)
1454.555 Million cell updates/sec

Title: US-10-070-277-2

Sequence: 1 MEISITQPDMDHLRLRDGDV.....SGDIIPFAGEMLDWLPAPL 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database:

1: A_Geneseq_16Dec04:*
2: geneseqp1380s:*
3: geneseqp1990s:*
4: geneseqp2000s:*
5: geneseqp2001s:*
6: geneseqp2002s:*
7: geneseqp2003a:*
8: geneseqp2003b:*
9: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1812	100.0	346	4	AAB70773
2	1812	100.0	346	4	AAB70779
3	1528	84.3	345	5	AAO15493
4	1067.5	58.9	364	6	ABU20245
5	1058	58.4	352	6	ABU22502
6	1039.5	57.4	350	6	ABU49750
7	1008.5	55.7	354	6	ABU23462
8	983	54.2	344	6	ABU37887
9	979	54.0	344	6	ABP78043
10	979	54.0	344	6	ABU37314
11	978	54.0	348	6	ABU31804
12	978	54.0	352	7	ABO61072
13	975	53.8	348	6	ABU38565
14	975	53.8	348	7	ABO68745
15	971	53.6	348	6	ABU15296
16	962	53.1	348	6	ABU39671
17	959	52.9	347	6	ABU42003
18	956	52.8	348	6	ABU50480
19	952.5	52.6	344	6	ABU16776
20	952.5	52.6	346	6	ADA34254
21	949	52.4	348	6	ABU47954
22	946	52.2	355	8	ADU04644
23	943	52.0	354	6	ABU35190
24	935	51.6	351	6	ABW68113
25	934	51.5	348	6	ABU44828

26	930	51.3	356	7	ADF04000	Adf04000 Bacterial
27	559	31.4	335	6	ABU26343	Abu26343 Protein e
28	541	29.9	339	6	ABU30820	Abu30820 Protein e
29	539.5	29.8	190	6	ABU27626	Abu27626 Protein e
30	527	29.1	371	2	AAW20981	Aaw20981 H. pylori
31	402	22.2	136	6	ABU40992	Abu40992 Protein e
32	350.5	19.3	116	5	ABP06300	Abp06300 Human ORF
33	209.5	11.6	146	2	AAW20558	Aaw20558 H. pylori
34	144	7.9	396	6	ABU24485	Abu24485 Streptococ
35	134.5	7.4	438	5	ABP26430	Abp26430 Streptococ
36	131.5	7.3	426	5	ABBA4703	Abba4703 Listeria
37	131.5	7.3	426	6	ABU32919	Abu32919 Protein e
38	128	7.1	424	6	ABU24236	Abu24236 Protein e
39	126.5	7.0	409	8	ADNA47121	Adna47121 Thermococ
40	123	6.8	397	4	ADBI1184	Adbi1184 Altolococ
41	121.5	6.7	427	6	AAU35265	Aau35265 Enterococ
42	121.5	6.7	427	6	ABU29149	Abu29149 Lactococc
43	120	6.6	441	5	ABBS4385	Abbs4385 Lactococc
44	119.5	6.6	426	6	ABU29702	Abu29702 Protein e
45	119.5	6.6	430	7	ADC95480	Adc95480 E. faeciu

ALIGNMENTS

RESULT 1
AAB70773 standard; protein; 346 AA.
ID AAB70773:
XX
AC AAB70773:
XX
DT 25-MAY-2001 (first entry)
XX
DE Potato dihydro-oxotase protein.
XX
KM Potato; dihydro-oxotase; plant; polysaccharide content; orotic acid;
KW uridine; starch synthesis.
XX
OS Solanum tuberosum.
XX
PN W0200114569-A2.
XX
PD 01-MAR-2001.
XX
PF 12-AUG-2000; 2000MO-EP007884.
XX
PR 20-AUG-1999; 99DE-01039688.
XX

(BADI) BASF AG.
Ehrhardt T, Stitt Nigel M, Geigenberger PL, Loeff I, Zrenner R;
Schroeder M;
WPI: 2001-202938/20.
DR N-PSDB; AAF61217.
XX
PT Increasing the polysaccharide, especially starch, content of plants, by
transforming with a sequence encoding dihydro-oxotase.
XX
PS Example 4; Page 30-31; 32pp; German.

This invention describes a novel use of a DNA sequence (I) encoding
dihydro-oxotase (II) to prepare plants which have increased
polysaccharide (PS) content. The invention also describes (1) a 1271 base
pair DNA sequence (I) encoding potato (II), sequences that hybridize to
it and its homologs (or their fragments); (2) method for preparing plants
of increased PS content by expressing (I) in them; (3) method for
transforming plants by introducing an expression cassette, containing
promoter, signal sequence and (I), into a plant cell, callus tissue,
complete plant or protoplast; and (4) plants with increased PS content
containing the DNA of (1). Expression of (II) increases the content of
orotic acid (a precursor of uridine nucleotides), resulting in an
increased rate of starch synthesis. Typically, in potato tuber slices,

CC addition of 10 mM orotate more than doubled the rate of starch synthesis.
 CC (1) is used to prepare plants (especially tomato, tobacco, potato,
 CC lipoic, manioc, rice, barley, oats, rye, wheat and maize) with increased
 CC PS, particularly starch, content
 XX Sequence 346 AA;

Query Match 100.0%; Score 1812; DB 4; Length 346;
 Best Local Similarity 100.0%; Pred. No. 1.6e-173;
 Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELSTTQDDMHLRLDGDVLAQVSHSAHFGRAIYMPKRPITTTAAVAAYREALIK 60
 DB 1 MELSTTQDDMHLRLDGDVLAQVSHSAHFGRAIYMPKRPITTTAAVAAYREALIK 60
 QY 61 SLVPDSDPNPLMTLYLDTTSPMEIKLARSQVVFVKLYPAGATNDSQDGTDLFGKCL 120
 DB 61 SLVPDSDPNPLMTLYLDTTSPMEIKLARSQVVFVKLYPAGATNDSQDGTDLFGKCL 120
 QY 121 PVLQEMVEHNMPLLVHGEVTNPEDVDFREKVFIEVLRPLVQKRPOLKVMHEHTTTIDA 180
 DB 121 PVLQEMVEHNMPLLVHGEVTNPEDVDFREKVFIEVLRPLVQKRPOLKVMHEHTTTIDA 180
 QY 181 VKFVESCTEGFVAATVTPQHLVLNNSLFGGLQPHNYCLPVLKREIHREALVSAVTSGS 240
 DB 181 VKFVESCTEGFVAATVTPQHLVLNNSLFGGLQPHNYCLPVLKREIHREALVSAVTSGS 240
 QY 241 KRFFLGTDSAPHDRRKECCSCCAGIYNAPVALSYAVFEKENALDLEAFTSFGNDPF 300
 DB 241 KRFFLGTDSAPHDRRKECCSCCAGIYNAPVALSYAVFEKENALDLEAFTSFGNDPF 300
 QY 301 YGLPRNNSKIKLSKTPMKVPESFYASGDIIPMFAEMLDWLPAPL 346
 DB 301 YGLPRNNSKIKLSKTPMKVPESFYASGDIIPMFAEMLDWLPAPL 346

RESULT 2
 AAB70779
 ID AAB70779 standard; protein; 346 AA.

AC AAB70779;
 XX
 DT 25-MAY-2001 (first entry)
 XX
 DE Potato dihydro-orotate protein.
 XX
 KM Potato; dihydro-orotate; inhibitor; herbicide; transgenic plant.
 XX
 OS Solanum tuberosum.
 XX
 FN WO200118190-A2.
 XX
 PD 15-MAR-2001.
 XX
 PF 02-SEP-2000; 2000WO-EP008581.
 XX
 PR 07-SEP-1999; 99DE-01042742.
 XX
 PA (BADI) BASF AG.
 XX
 PI Ehrhardt T, Lerchl J, Stilt Nigel M, Zrenner R, Schroeder M,
 XX
 DR WPI; 2001-235198/24.
 XX
 DR N-PSDB; AAF61262.
 XX

PT New DNA encoding plant dihydroorotate, for producing herbicide-resistant
 PT plants and for screening for compounds that can inhibit dihydroorotate
 PT and that be used as herbicides.
 XX

PS Claim 3; Page 31-32; 38pp; German.
 XX

CC This invention describes a novel DNA sequence (1) containing the coding
 CC region for a plant dihydro-orotate (DHO), comprising a sequence (S1) of

CC 1271 base pairs (bp), given in the specification. The invention also
 CC describes (1) a DNA sequence (1a) that hybridizes to (S1), or its parts
 CC or derivatives produced by insertion, deletion or substitution, and
 CC encodes a protein with DHO activity; (2) a protein (11) containing at
 CC least 100 amino acids (aa) from a 346 aa sequence (S2), given in the
 CC specification; (3) identifying substances (111) that inhibit activity of
 CC plant DHO; (4) identifying substances (111a) which act as herbicides by
 CC inhibition of plant DHO; (5) a test system based on expression of (S1)
 CC for identifying herbicidal inhibitors of DHO; and (6) inhibitors of plant
 CC DHO identified by the system of (5). The products of the invention have
 CC herbicide activity. (1) is used: (i) for recombinant production of DHO in
 CC prokaryotic or eukaryotic cells; (ii) to produce a test system for
 CC identifying compounds that inhibit DHO, potentially useful as herbicides;
 CC and (iii) to produce transgenic plants that have increased resistance to
 CC DHO-inactivating herbicides. Recombinantly expressed DHO provides a
 CC reproducible and rapid (high throughput) screen for specific inhibitors
 CC of activity
 XX
 XX Sequence 346 AA;

Query Match 100.0%; Score 1812; DB 4; Length 346;
 Best Local Similarity 100.0%; Pred. No. 1.6e-173;
 Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELSTTQDDMHLRLDGDVLAQVSHSAHFGRAIYMPKRPITTTAAVAAYREALIK 60
 DB 1 MELSTTQDDMHLRLDGDVLAQVSHSAHFGRAIYMPKRPITTTAAVAAYREALIK 60
 QY 61 SLVPDSDPNPLMTLYLDTTSPMEIKLARSQVVFVKLYPAGATNDSQDGTDLFGKCL 120
 DB 61 SLVPDSDPNPLMTLYLDTTSPMEIKLARSQVVFVKLYPAGATNDSQDGTDLFGKCL 120
 QY 121 PVLQEMVEHNMPLLVHGEVTNPEDVDFREKVFIEVLRPLVQKRPOLKVMHEHTTTIDA 180
 DB 121 PVLQEMVEHNMPLLVHGEVTNPEDVDFREKVFIEVLRPLVQKRPOLKVMHEHTTTIDA 180
 QY 181 VKFVESCTEGFVAATVTPQHLVLNNSLFGGLQPHNYCLPVLKREIHREALVSAVTSGS 240
 DB 181 VKFVESCTEGFVAATVTPQHLVLNNSLFGGLQPHNYCLPVLKREIHREALVSAVTSGS 240
 QY 241 KRFFLGTDSAPHDRRKECCSCCAGIYNAPVALSYAVFEKENALDLEAFTSFGNDPF 300
 DB 241 KRFFLGTDSAPHDRRKECCSCCAGIYNAPVALSYAVFEKENALDLEAFTSFGNDPF 300
 QY 301 YGLPRNNSKIKLSKTPMKVPESFYASGDIIPMFAEMLDWLPAPL 346
 DB 301 YGLPRNNSKIKLSKTPMKVPESFYASGDIIPMFAEMLDWLPAPL 346

RESULT 3
 AA015493
 ID AA015493 standard; protein; 345 AA.

AC AA015493;
 XX
 DT 03-OCT-2002 (first entry)
 XX
 DE Beta vulgaris dihydroorotate.
 XX
 KM Sugar beet; transgenic plant; osmotic stress tolerance;
 KM oxidative stress tolerance; casein kinase alpha catalytic subunit;
 KM dihydroorotate; translation initiation factor 1a; plant flowering;
 KM plant growth stimulation.
 XX
 OS Beta vulgaris.
 XX
 PN WO200252012-A2.
 XX
 PD 04-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-EP015093.
 XX

PR 22-DEC-2000; 2000EP-00870319.
PR 26-FEB-2001; 2001US-0271656P.
XX
PA (CROP-) CROPEDESIGN NV.
PI Kanhonou RA, Serrano Salom R, Ros Palau R;
XX WPI; 2002-583567/62.
DR N-PSDB; AAL44159.
XX
PT Novel sugar beet genes involved in stress response, useful for rendering
PT crops resistant to stress situations like osmotic stress caused by salt,
XX drought, cold or frost thus increasing crop yield.
XX
PS Claim 26; Fig 6; 95pp; English.
XX
CC The invention comprises the amino acid and coding sequences of sugar beet
CC (Beta vulgaris) genes that are involved in osmotic and oxidative stress
CC tolerance in a plant. The Beta vulgaris genes of the invention encode the
CC proteins: casein kinase alpha catalytic subunit; dihydroorotase;
CC transactin initiation factor 1a; putative protein; and an unknown
CC protein. The Beta vulgaris genes of the invention are useful for
CC enhancing the osmotic and oxidative stress tolerance of a plant. The Beta
CC vulgaris genes are also useful for stimulating plant growth. The casein
CC kinase alpha catalytic subunit gene is useful for controlling the process
CC of flowering in a plant. The present amino acid sequence represents the
CC Beta vulgaris dihydroorotase
XX
SQ Sequence 345 AA;
Query Match 84.3%; Score 1528; DB 5; Length 345;
Best Local Similarity 82.0%; Pred. No. 6.9e-145;
Matches 282; Conservative 34; Mismatches 28; Indels 0; Gaps 0;
QY 1 MEISTQPDMMHLIRGCVLKAIVSHSAHFGRAIWMPLKPRITTTAAAVYRAILK 60
DB 1 MELTLRPMWHLIRGCVLKAIVSHSAHFGRAIWMPLKPRITTTAAAVYRAILK 60
QY 61 SLVPDSDFNPLMTLYLTDITTSPEIKLARSQVFGVLYPAGATINSODGVTDLFGKCL 120
DB 61 VLPSDSDFNPLMTLYLTDITTSPEIKLARSQVFGVLYPAGATINSODGVTDLFGKCL 120
QY 121 PVLQEWVHNHMLLVHGEVTNPEVDMFDEKRVIFETVLRLVQKFPOLKVMMEHTTTDA 180
DB 121 PVLEEMABQDMPLLVHGEVTNPEVDMFDEKRVIFETVLRLVQKFPOLKVMMEHTTTDA 180
QY 181 VKPEFESCTEGVNAATVPQHLVLRNLSLFOGGLQPHNVCPLVKREIHRALVSATVSGS 240
DB 181 VKFIESCNGGVNAATVPQHLVLRNLSLFOGGLQPHNVCPLVKREIHRALVSATVSGS 240
QY 241 KRFLIGTDSAPHDRRRKSCGCGAGIYVAPALSVYAKFEKENALDKLEAFTSPNGPDF 300
DB 241 KQYLGTDSDAPHRRRKSCGCGAGIYVAPALSVYAKFEKENALDKLEAFTSPNGPDF 300
QY 301 YGLPRNNSKIKLSTKTPKVPESFSYASGDIIPMFAGKMLDWLPA 344
DB 301 YGLPRNNSKIKLSTKTPKVPESFSYASGDIIPMFAGKMLDWLPA 344
RESULT 4
ABU20245
ID ABU20245 standard; protein; 364 AA.
XX
AC ABU20245;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #5772.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS Borrelia cepacia.

PN MO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342823P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
PI Wall D, Trawick JD, Carr GD, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA24115.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 48169; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-regulated gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 364 AA;
Query Match 58.9%; Score 1067.5; DB 6; Length 364;
Best Local Similarity 60.4%; Pred. No. 1.9e-98;
Matches 206; Conservative 37; Mismatches 95; Indels 3; Gaps 2;
QY 3 LSTQPDMMHLIRGCVLKAIVSHSAHFGRAIWMPLKPRITTTAAAVYRAILKSL 62
DB 22 LSLARPDMMHLIRGCVLKAIVSHSAHFGRAIWMPLKPRITTTAAAVYRAILKSL 62
QY 63 PVDSDFNPLMTLYLTDITTSPEIKLARSQVFGVLYPAGATINSODGVTDLFGKCLPV 122
DB 82 PAGMTFEPMLTLYLTDITTSPEIKLARSQVFGVLYPAGATINSODGVTDLFGKCLPV 140
QY 123 LQEWVHNHMLLVHGEVTNPEVDMFDEKRVIFETVLRLVQKFPOLKVMMEHTTTDAVK 182

```

Db      141 LEAMQETGMPPLVHGSVTDASIDLFDRREKVFIDVMEFLRDFGLKAVFEHITTKAD 200
Qy      183 FVE--SCTEGFVAATVTPQHLVLRNSLFGGLOPHNYCLPVKREIHRALVSAVTSGS 240
Db      201 YVRADAAAPGLIGATITTAHLLVLRNMLFVGCIIRPHYCLPVLRERHVALVEAATSGN 260
Qy      241 KRFFLGTDSPAPDRRKSCGCGAGIYNAPVALSVAKVEKENALDKLEAFSTNGPDF 300
Db      261 PRFFLGTDSPAPDRRKSCGCGAGIYNAPVALSVAKVEKENALDKLEAFSTNGPDF 320
Qy      301 YGLPRNSKIKLSKTTPKVPESFSYASGDIIMPAGEMLDW 341
Db      321 YGLPRSAETVTLRREPMELPREIRPAGETPVVPLRGSETIGW 361

```

RESULT 5
ABU22502
ID ABU22502 standard; protein; 352 AA.
XX
AC ABU22502;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by prokaryotic essential gene #8029.
XX
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Burkholderia mallei.
XX
FN W0200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.
XX
PR 08-FEB-2002; 2002US-00072851.
XX
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.

XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KI, Zykkind JW;
PI Wall D, Trivick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
XX
DR N-PSDB; ACA26372.
XX
XX

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX

PS Claim 25; SEQ ID NO 50426; 1766bp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC from wipo.int/pub/published_pct_sequences
XX

Seq Sequence 352 AA;
Query Match 58.4%; Score 1058; DB 6; Length 352;
Best Local Similarity 58.4%; Pred. No. 1.7e-97;
Matches 199; Conservative 45; Mismatches 95; Indels 2; Gaps 1;

```

Qy      3 LSTIQPDWHLRLDRDVLKAVSHSAHFGRAIWMNLKPPITTTAAVAVERAILKSL 62
Db      9 LTLARPDDWHLVHVDGMLAVALPHTARQGRBALIMNLPVTTTHQAQVERILAAV 68
Qy      63 PVDSDPNPLMTLYLTPTTSPWEIKLARESOVFGVKLYPAGATTNSQDGVTDLFGCLPV 122
Db      69 PACMTFFELMTLYLTDTTPADEIRRARESCGVGKLYPAGATTNSDAGVTDLGGCART 128
Qy      123 LOENVEHNMPLVHGEVTNPEVDMFDEKVFIEFTVLRPIVQKPOLKVMHEVTTIDAVK 182
Db      129 LEAMQEVGMPLVHGEVTDPSIDLFRREKVFIDVMEFLRDLGLKAVFEHITTKAD 188
Qy      183 FVE--SCTEGFVAATVTPQHLVLRNSLFGGLOPHNYCLPVKREIHRALVSAVTSGS 240
Db      189 YVRADAAASRGATITTAHLLVLRNMLFVGCIIRPHYCLPVLRERHVALVEAATSGN 248
Qy      241 KRFFLGTDSPAPDRRKSCGCGAGIYNAPVALSVAKVEKENALDKLEAFSTNGPDF 300
Db      249 PRFFLGTDSPAPDRRKSCGCGAGIYNAPVALSVAKVEKENALDKLEAFSTNGPDF 308
Qy      301 YGLPRNSKIKLSKTTPKVPESFSYASGDIIMPAGEMLDW 341
Db      309 YGLPRSAETVTLRREPMELPREIRPAGETPVVPLRGSETIGW 349

```

RESULT 6
ABU49750
ID ABU49750 standard; protein; 350 AA.
XX
AC ABU49750;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by prokaryotic essential gene #35277.
XX
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Vibrio cholerae.
XX
FN W0200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.
XX
PR 08-FEB-2002; 2002US-00072851.
XX
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haasebeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI: 2003-029926/02.
DR N-PSDB; ACA53620.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 77674; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway;
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 350 AA;
Query Match 57.4%; Score 1039.5; DB 6; Length 350;
Best Local Similarity 59.0%; Pred. No. 1.2e-95;
Matches 200; Conservative 45; Mismatches 91; Indels 3; Gaps 3;

QY 3 LSTGPRDWHHLRGGDVLKAVSHHPRGATVMPKLPITTAANAAYREALTSL 62
DB LTRPRDWHHLRGGDVLKAVSHHPRGATVMPKLPITTAANAAYREALTSL 71
QY 63 PVDSPFNPLMTLYLDTTSPMEIKLARSOVFGVKLPAGATTNSODGVTLDFKCLPV 122
DB 72 P-QAHFELMALYLTDNTSPERIRAKXSGKVAAKLVPAGATTNSDGSVTS-A-KNIPV 129
QY 123 LOEVENHMPLLVHSEVTNPEDVDFREKVFETVLRPLVQKFPOLKAVMEHTTIDAVK 182
DB 130 LQAMQEVGMLLVHSEVTNPEDVDFREKVFETVLRPLVQKFPOLKAVMEHTTIDAVK 189
QY 183 FVESTGESFVATVTPHVLVNRNSLFGGLOPHNYCPVLKRETHREALVSAVTSQSKR 242
DB 190 FVQOAGDN-VAAITTAHHLFERNMVLVGGIRPHFYCAPILKRAVTHQALVAATSSGSK 248
QY 243 FFLGTDSAPHDRRRKSCGCGAGIYAPVALSVKAVEKENALDKLEATFSFNGPDYFG 302
DB 249 FFLGTDSAPHAKGREAAACGAGSTYAAALVLAVEFKEGKLELAFSFGNPDYFG 308
QY 303 LPRNNSKITLSTKTPMKVPSFSYVAGDITPMFAGEMLDW 341

DB 309 LPRNQETVTLTKQAMPVASEMPGSDIVVPIRAGENIEW 347
RESULT 7
ABU23462
ID ABU23462 standard; protein; 354 AA.
XX
AC ABU23462;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #9989.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS *Bordetella pertussis*.
XX
PN MO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002MO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362659P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haasebeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI: 2003-029926/02.
DR N-PSDB; ACA27332.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 51386; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway;
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 354 AA;

Query Match 55.7%; Score 1008.5; DB 6; Length 354;
Best Local Similarity 57.0%; Pred. No. 1.6e-92;
Matches 196; Conservative 46; Mismatches 97; Indels 5; Gaps 3;

```
QY 2 ELSTQPDMDHLNRDGDVLAQAVSHSAHFGRAIVMPNKKPITTTAAVAVR--EAI 58
DB 8 ELTTRPDMDHLNRDGDVLAQAVSHSAHFGRAIVMPNKKPITTTAAVAVR--EAI 58
QY 59 LKSPVQ-SDFNPMTLYITDTSPEMEIKLARSQVFGVLYPAGATTNSODGVTDLEFG 117
DB 68 LKRGSGTAAFTPLMTLYITDNTPAEIVAHESGQVAVKLYPGATTNSAGTTDLG 127
QY 118 KCLPVLQGVNHNMPVLHGEVTPNPEVMDREKVEIEVLAIVQKFPOLKVMHEVTT 177
DB 128 KCGAALAALERCGMPILVHGEVTPDAIVDFDEAVFIERVMQPLRRAYPGLKVFEEHITT 187
QY 178 IDAVKFESECTEGFAATVTPQHLVINRNSLFOGGLQPHNYCLPVLEKEIHREALVASVT 237
DB 188 REGAHYVRD-ASGPTATITTPQHMLYNRNAIFTGVRPHWCLPILKEVHRQLVEMAT 246
QY 238 SSGSKPEFLGTDNAPHRRRKSCSCGAGIYNAPVALSVYAKPEKENALDTLEAFSPNG 297
DB 247 SSGSPREFLGTDNAPHRRGKKEHACGACCTYLAHMELYATAFPAVGRLDRLBGFASFHG 306
QY 298 PDVYGLPRNNSKIKLSKTPWKPESFSYASGDIIPMFAGEMLDW 341
DB 307 PDVYGLPRNNTGTLTTRREAVETPAEVAFGDTLVVPLSGSESLGW 350
```

RESULT 8
ABU37887
ID ABU37887 standard; protein; 344 AA.
XX
AC ABU37887;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #23414.

KW Antisense: prokaryotic essential gene; cell proliferation; drug design.

OS Neisseria meningitidis.

BN WO200277183-A2.

PD 03-OCT-2002.

PF 21-MAR-2002; 2002MO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR N-PSDB; ACA11757.

DR MPI; 2003-029926/02.

PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 65811; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

Sequence 344 AA;

Query Match 54.2%; Score 983; DB 6; Length 344;
Best Local Similarity 56.3%; Pred. No. 5.8e-90;
Matches 192; Conservative 55; Mismatches 88; Indels 6; Gaps 5;

```
QY 3 LSTQPDMDHLNRDGDVLAQAVSHSAHFGRAIVMPNKKPITTTAAVAVREALKSL 62
DB 4 LTRTPDMDHLNRDGDVLAQAVSHSAHFGRAIVMPNKKPITTTAAVAVREALKSL 62
QY 63 PVDSFNPMTLYITDTSPEMEIKLARSQVFGVLYPAGATTNSODGVTDLEFGCLPV 122
DB 64 PEGSAFEPMTLYITDTPALVREKAAQIV-AKLTLPAGATTNSODGVTDLEF-KLIPV 121
QY 123 LQEMVHNMPVLHGEVTPNPEVMDREKVEIEVLAIVQKFPOLKVMHEVTTIDAVK 182
DB 122 LEMNAKQGLTVHGEVTPNPEVMDREKVEIEVLAIVQKFPOLKVMHEVTTIDAVK 181
QY 183 FVESCTEGFAATVTPQHLVINRNSLFOGGLQPHNYCLPVLEKEIHREALVASVTSSK- 241
DB 182 LVLEAGDN-VAAFTVPGHLLNRNDLVGVGRPHFCPLVAKRTTHQALVAAT-GEKA 239
QY 242 -RFFLGTSAPHRDRKSCSCGAGIYNAPVALSVYAKPEKENALDTLEAFSPNGDP 300
DB 240 HRFELGTSAPHRDRKSCSCGAGIYNAPVALSVYAKPEKENALDTLEAFSPNGDP 299
QY 301 YGLPRNNSKIKLSKTPWKPESFSYASGDIIPMFAGEMLDW 341
DB 300 YGIPENTDTITLVKQSGTVPASIFYGDELVPWBAAGEIGW 340
```

RESULT 9
ABP78043
ID ABP78043 standard; protein; 344 AA.
XX
AC ABP78043;

DT 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 2616.

KW Antibacterial; infection; vaccine; gene therapy.
OS Neisseria gonorrhoeae.
XX MO200279243-A2.
XX
XX 10-OCT-2002.
XX
XX 12-FEB-2002; 2002WO-IB002069.
XX
XX 12-FEB-2001; 2001GB-00003424.
XX
XX (CHIR-) CHIRON SPA.
XX Fontana MR, Piazza M, Maignani V, Monaci E;
XX WPI; 2003-058415/05.
XX N-PSDB; ABZ39013.
XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX medicament for treating or preventing N. gonorrhoeae infection.
XX
XX Disclosure; Page 379; 815pp; English.
XX
XX The present invention relates to proteins from Neisseria gonorrhoeae.
XX Also disclosed are the nucleic acid molecules encoding the proteins and
XX antibodies that specifically bind to the proteins. The composition
XX comprising the protein, nucleic acid or antibody is useful for the
XX manufacture of a medicament for treating or preventing N. gonorrhoeae
XX infection, this may be in the form of a vaccine or gene therapy.
XX Sequences given in records ABZ6736-ABZ61046 represent nucleic acid
XX molecules of the invention
XX
XX Sequence 344 AA;
SQ
Query Match 54.0%; Score 979; DB 6; Length 344;
Best Local Similarity 56.3%; Pred. No. 1.5e-89;
Matches 192; Conservative 55; Mismatches 88; Indels 6; Gaps 5;
QY 3 LSTQPDHMLHLDGDLKAVVSHHFGRAIVMPKLPITTTAAVAYRAILKSL 62
DB 4 LTIIRPDHMLHLDGDLKAVVSHHFGRAIVMPKLPITTTAAVAYRAILKSL 63
QY 63 PVDSDPMLTLVLTDTTSPMEIKLARESQVFGVKLYPAGATNSODGVTDLFGKCLPV 122
DB 64 PEGSAFERMLTLVLTDTTSPMEIKLARESQVFGVKLYPAGATNSODGVTDLFGKCLPV 121
QY 123 LQENVEHNMPLLVHGEVTPPEVDMFDRKQVIEIVLRPLVOKFQOLKYVMEHVTIDAVK 182
DB 122 LEEWAKQGIPLVHGEVTPPEVDMFDRKQVIEIVLRPLVOKFQOLKYVMEHVTIDAVK 181
QY 183 FVSECTSEFVAATVTPQHLVYNNRSLPGCGLOPHNYCLPVKREIHRFALVSAVTSQSK- 241
DB 182 LVLAGDND-VAATVTPQHLVYNNRSLPGCGLOPHNYCLPVKREIHRFALVSAVTSQSK- 239
QY 242 -RFFLGDTSDAPDRRRKSCGCGAGIYVAPALSVYAVFKEKNALDLEAFSTENGPDF 300
DB 240 HKFPLGDTSDAPDRRRKSCGCGAGIYVAPALSVYAVFKEKNALDLEAFSTENGPDF 299
QY 301 YGLPRNNSKITLSTKTPWKVPSFSYASGDIIPMFAGENLMDV 341
DB 300 YGIPENADTITLVKOSQTVPAVSPYGDGLVPMRAGGIGW 340
RESULT 10
ABU37314
ID ABU37314 standard; protein; 344 AA.
XX
XX AC ABU37314;
XX
XX 23-OCT-2003 (revised)
DT 19-JUN-2003 (first entry)
XX

DE Protein encoded by Prokaryotic essential gene #22841.
XX
XX Antibiotic; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Neisseria gonorrhoeae.
XX
XX MO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242-
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342823P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX
XX (BLIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA41184.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 65238; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation, or for screening homologous nucleic acids
XX drug discovery programs, or for screening homologous nucleic acids
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
XX standardise OS field)
XX
XX Sequence 344 AA;
SQ
Query Match 54.0%; Score 979; DB 6; Length 344;
Best Local Similarity 56.3%; Pred. No. 1.5e-89;
Matches 192; Conservative 55; Mismatches 88; Indels 6; Gaps 5;
QY 3 LSTQPDHMLHLDGDLKAVVSHHFGRAIVMPKLPITTTAAVAYRAILKSL 62
DB 4 LTIIRPDHMLHLDGDLKAVVSHHFGRAIVMPKLPITTTAAVAYRAILKSL 63
QY 63 PVDSDPMLTLVLTDTTSPMEIKLARESQVFGVKLYPAGATNSODGVTDLFGKCLPV 122
DB 64 PEGSAFERMLTLVLTDTTSPMEIKLARESQVFGVKLYPAGATNSODGVTDLFGKCLPV 121
QY 123 LQENVEHNMPLLVHGEVTPPEVDMFDRKQVIEIVLRPLVOKFQOLKYVMEHVTIDAVK 182
DB 122 LEEWAKQGIPLVHGEVTPPEVDMFDRKQVIEIVLRPLVOKFQOLKYVMEHVTIDAVK 181
QY 183 FVSECTSEFVAATVTPQHLVYNNRSLPGCGLOPHNYCLPVKREIHRFALVSAVTSQSK- 241
DB 182 LVLAGDND-VAATVTPQHLVYNNRSLPGCGLOPHNYCLPVKREIHRFALVSAVTSQSK- 239
QY 242 -RFFLGDTSDAPDRRRKSCGCGAGIYVAPALSVYAVFKEKNALDLEAFSTENGPDF 300
DB 240 HKFPLGDTSDAPDRRRKSCGCGAGIYVAPALSVYAVFKEKNALDLEAFSTENGPDF 299
QY 301 YGLPRNNSKITLSTKTPWKVPSFSYASGDIIPMFAGENLMDV 341
DB 300 YGIPENADTITLVKOSQTVPAVSPYGDGLVPMRAGGIGW 340

Db 4 LTIIRPDDMHILNRDGDALKAVALPYTARQWGRAVIMPNLKPVSVAADALVAKATIMATL 63
QY 63 PVSDPFPNPLMTLYLTDTTSPMEIKLARSEQVYGVKLYPAGATTNSQDGVTDLPFGKCLPV 122
Db 64 PESSAPPEPLMTLYLTDTTSPMEIKLARSEQVYGVKLYPAGATTNSQDGVTDLPFGKCLPV 122
QY 123 LOEMVENHNPPLVHGEVTNPEVDMFDEKVFIEITVLRPLVQKFPOLKVMHEHTTIDAVK 182
Db 122 LEMNAQGLIFLVHGEVTDPEIDIFDEKVFIEITVLRPLVQKFPOLKVMHEHTTIDAVK 181
QY 183 FVESCIEGFVAATVTPPOHVTNPNLSFOGLOPHNYCLPVLKREITHREALVSAVTSGRK- 241
Db 182 LVLEAGDN-VAAITVTPPOHVTNPNLSFOGLOPHNYCLPVLKREITHREALVSAVTSGRK- 241
QY 242 -RFFLTGDSAPDRRRKESCCGAGIYNAPVALSYAVKFEKENALDKLEAFTSPNGPDF 300
Db 240 HKFFLTGDSAPDRRRKESCCGAGIYNAPVALSYAVKFEKENALDKLEAFTSPNGPDF 300
QY 301 YGLPNNNSKIKLSTKTPMKVPSEFSYASGDIIPMFAGEMLDW 341
Db 300 YGIPENADTTTLVKOSQTVPAVSPYGDGLVPMRAGEIGW 340

RESULT 11

ABU31804

ID ABU31804 standard; protein; 348 AA.

XX AC ABU31804;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #17331.

XX KM Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Klebsiella pneumoniae.

XX PN W020027183-A2.

XX PD 03-OCT-2002.

XX PE 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KJ, Zyckind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA35674.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 59728; 1766bp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for

CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC product's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 348 AA;

XX SQ Query Match 54.0%; Score 978; DB 6; Length 348;

XX SQ Best Local Similarity 55.8%; Pred. No. 1.9e-89;

XX SQ Matches 189; Conservative 46; Mismatches 102; Indels 2; Gaps 2;

QY 3 LTIIRPDDMHILNRDGDALKAVALPYTARQWGRAVIMPNLKPVSVAADALVAKATIMATL 62
Db 8 LTIIRPDDMHILNRDGDALKAVALPYTARQWGRAVIMPNLKPVSVAADALVAKATIMATL 62
QY 63 PVSDPFPNPLMTLYLTDTTSPMEIKLARSEQVYGVKLYPAGATTNSQDGVTDLPFGKCLPV 122
Db 68 PVSDPFPNPLMTLYLTDTTSPMEIKLARSEQVYGVKLYPAGATTNSQDGVTDLPFGKCLPV 122
QY 123 LOEMVENHNPPLVHGEVTNPEVDMFDEKVFIEITVLRPLVQKFPOLKVMHEHTTIDAVK 182
Db 127 LOEMVENHNPPLVHGEVTNPEVDMFDEKVFIEITVLRPLVQKFPOLKVMHEHTTIDAVK 182
QY 183 FVESCIEGFVAATVTPPOHVTNPNLSFOGLOPHNYCLPVLKREITHREALVSAVTSGRK 241
Db 187 FVESCIEGFVAATVTPPOHVTNPNLSFOGLOPHNYCLPVLKREITHREALVSAVTSGRK 241
QY 242 -RFFLTGDSAPDRRRKESCCGAGIYNAPVALSYAVKFEKENALDKLEAFTSPNGPDF 300
Db 246 -RFFLTGDSAPDRRRKESCCGAGIYNAPVALSYAVKFEKENALDKLEAFTSPNGPDF 300
QY 301 YGLPNNNSKIKLSTKTPMKVPSEFSYASGDIIPMFAGEMLDW 341
Db 306 YGLPNNNSKIKLSTKTPMKVPSEFSYASGDIIPMFAGEMLDW 341

RESULT 12

ID ABO61072 standard; protein; 352 AA.

XX AC ABO61072;

XX DT 29-JUL-2004 (first entry)

XX DE Klebsiella pneumoniae polypeptide seqid 7589.

XX KM Recombinant expression vector; transcription regulatory element;

XX KM Klebsiella pneumoniae protein; antibacterial; Vaccine.

XX OS Klebsiella pneumoniae.

XX PN US610836-B1.

XX PD 26-AUG-2003.

XX PF 27-JAN-2000; 2000US-00489039.

XX 29-JAN-1999; 99US-0117747P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX Breton GL, Osborne M;
XX WPI; 2003-895346/82.
XX N-PSDB; ACH94623.
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
XX preparing a vaccine composition against Klebsiella pneumoniae.
XX Disclosure; SEQ ID NO 7589; 932bp; English.
XX The invention describes a new isolated nucleic acid encoding a Klebsiella
XX pneumoniae polypeptide. Also described are: a recombinant expression
XX vector comprising the nucleic acid, operably linked to a transcription
XX regulatory element; and a cell comprising the recombinant expression
XX vector. The nucleic acid is useful for preparing a vaccine composition
XX against Klebsiella pneumoniae. This is the amino acid sequence of a
XX Klebsiella pneumoniae polypeptide of the invention
XX
SQ Sequence 352 AA:
Query Match 54.0%; Score 978; DB 7; Length 352;
Best Local Similarity 55.8%; Pred. No. 1.9e-89;
Matches 189; Conservative 46; Mismatches 102; Indels 2; Gaps 2;
QY 3 LSTQPPDMLHLDGQVLAAYVSHSAHNGRAIVMNLKPPITTTAAAVYREAILKSL 62
DB 12 LKTRRPDMDHILHDDMDLKTVPYTSFGRATVMPNVPVTVAAAIAYRORIMDAV 71
QY 63 PVSDPFLMTLYLTDTTSPMEIKLARESQVFGVKLYPGATTNSQDVTDLFGKCLPV 122
DB 72 PAGHDFPLMTCYITDSLDPAELERGNREGVFTAAKIYPANATNSSHGVTST-DALMPV 130
QY 123 LQENVEHNMPLLVHGEVNTPEVDMFDRKQVFIETVLBPVOKFPOLKVMHEVTTDAVK 182
DB 131 LERNEKLGMPLLVHGEVTHAIDIFDRARFIETVMEPLRQLGLKVFEEHITTKDAAE 190
QY 183 FVESCGRFVAATVTPQHLVLRNSLFOGGLQPHNYCLPVLKREIHEALVSAVTSQSKR 242
DB 191 YVRGNGE-LLAATTPQHLMFNRNMMLVGGIRPHLYCLPVLRKNIHQALRELVASGFSR 249
QY 243 PFLGTDSAPHDRRKRECGCGAGIYNAPVALSVYAKYPEKENALDKLEAFTSFNGPDYFG 302
DB 250 AFLGIDSLFPHARKHKEKSCGACGCPNAPTALGSTAYFEEMNALQHEPAFCSLNGPRFYG 309
QY 303 LPRNNSKIKLSKTPMKYPESESFVSAGDIIPMFAGEMLDW 341
DB 310 LPVNESYVELVREETTVDSIALPNDTLVPFLAGSTVW 348
RESULT 13
ABU38565
ID ABU38565 standard; protein; 348 AA.
XX
AC ABU38565;
XX
XX 19-JUN-2003 (first entry)
XX DT
XX DE Protein encoded by Prokaryotic essential gene #24092.
XX DE
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Pseudomonas aeruginosa.
XX PN WO20027183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 05-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haseelbeck R, Ohleen KU, Zyckind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA42435.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 66489; 1766bp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-regulated gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
XX patent did not form part of the printed specification, but was obtained
XX in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 348 AA:
Query Match 53.8%; Score 975; DB 6; Length 348;
Best Local Similarity 55.5%; Pred. No. 3.8e-89;
Matches 188; Conservative 43; Mismatches 106; Indels 2; Gaps 2;
QY 3 LSTQPPDMLHLDGQVLAAYVSHSAHNGRAIVMNLKPPITTTAAAVYREAILKSL 62
DB 5 LTLRPDMDHILHLDGQALANTVGDARTGRATVMPNVPVTVAAAEADVQRRIILAR 64
QY 63 PVSDPFLMTLYLTDTTSPMEIKLARESQVFGVKLYPGATTNSQDVTDLFGKCLPV 122
DB 65 PAASRFPPLMWLYLTDRSTEEIRTAASGVHAAKIYPGATTNSQGVTRI-DNIFEA 123
QY 123 LQENVEHNMPLLVHGEVNTPEVDMFDRKQVFIETVLBPVOKFPOLKVMHEVTTDAVK 182
DB 124 LEAAAEVGMPLLVHGEVTVRAVDVDFDRKQFIDHRLRVERFPTLVKVFEEHITTKDAQ 183
QY 183 FVESCGRFVAATVTPQHLVLRNSLFOGGLQPHNYCLPVLKREIHEALVSAVTSQSKR 242

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 12, 2005, 05:25:37 ; Search time 102 Seconds
(without alignments)
1737.053 Million cell updates/sec

Title: US-10-070-277-2

Perfect score: 1812
Sequence: 1 MELSTITQDDMHLLHRDGV.....SGDIIMFAGEMLDMLPAFL 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1528	84.3	345	2 Q8VXS9	Q8VXS9 beta vulgar
2	1467	81.0	377	1 PYRC_ARATH	O04904 arabidopsis
3	1458	80.5	359	2 Q94J38	Q94J38 oryza sativ
4	1083	59.8	349	2 Q82WF3	Q82WF3 nitrosoma
5	1061	58.6	344	1 PYRC_RALSO	O8Y249 ralsoma s
6	1058	58.4	352	2 Q62H55	Q62H55 burkholderi
7	1058	58.4	352	2 Q63QW0	Q63QW0 burkholderi
8	1056.5	58.3	342	1 PYRC_SYNY3	P74438 synchocyst
9	1053	58.1	361	2 Q7NQ12	Q7NQ12 chromobacte
10	1039.5	57.4	342	1 PYRC_VIBCH	Q9K124 vibrio chol
11	1012.5	55.9	342	2 Q6LPI7	Q6LPI7 photobacter
12	1012.5	55.9	359	2 Q7M4A4	Q7M4A4 bordetella
13	1012.5	55.9	359	2 Q7WFR2	Q7WFR2 bordetella
14	1008.5	55.7	359	2 Q7VUP3	Q7VUP3 bordetella
15	1000	55.2	343	2 Q8B40	Q8B40 shewanella
16	991.5	54.7	342	1 PYRC_VIBPA	O87J46 vibrio para
17	986.5	54.4	347	2 Q7MF3	Q7MF3 vibrio vuln
18	983	54.2	344	1 PYRC_NEIMA	O9YV65 neisseria m
19	982	54.2	344	1 PYRC_NEIMA	O9K0D1 neisseria m
20	979.5	54.1	342	1 PYRC_VIBVU	O83C19 vibrio vuln
21	975	53.8	348	2 PYRC_PSEAE	P72170 pseudomonas
22	972	53.6	348	2 Q83RT8	Q83RT8 shigella fi
23	971	53.6	347	1 PYRC_ECOLI	P05020 escherichia
24	970	53.5	347	1 PYRC_ECOLI	O8X8N8 escherichia
25	967	53.4	347	2 Q6D658	Q6D658 erwinia car
26	966	53.3	347	1 PYRC_ECOLI	O8H1Z2 escherichia
27	962	53.1	347	1 PYRC_PSEBK	O8BNW7 pseudomonas
28	959	52.9	347	1 PYRC_PSEBM	O87XMI pseudomonas
29	955	52.8	348	1 PYRC_YERPE	O8FNU4 yersinia pe
30	955	52.7	348	2 Q669R2	Q669R2 yersinia ps
31	951	52.5	345	1 PYRC_AGRIS	O8U199 agrobacteri

32	951	52.5	347	1 PYRC_SALTY	P06204 salmonella
33	949	52.4	347	1 PYRC_SALTY	O8Z712 salmonella
34	939.5	51.8	344	2 Q6PD29	O6F429 actinobact
35	935	51.6	350	2 Q7NSW0	Q7NSW0 photorhabd
36	931	51.4	347	1 PYRC_RHIME	Q928C7 rhizobium m
37	925	51.0	351	2 Q83B17	Q83B17 coxiella bu
38	913	50.4	343	2 Q7U618	Q7U618 synchococc
39	901	49.7	349	2 Q7V2B2	Q7V2B2 prochlorococ
40	899.5	49.6	354	2 Q7VD18	Q7VD18 prochlorococ
41	893.5	49.3	354	2 Q7W896	Q7W896 prochlorococ
42	859.5	47.4	350	1 PYRC_BUCAI	P57416 buchnera ap
43	856.5	47.3	349	1 PYRC_BUCAP	O8K917 buchnera ap
44	697.5	38.5	349	2 Q8D3B4	Q8D3B4 wigglewort
45	596.5	32.9	204	1 PYRC_SERMA	Q98361 serratalia ma

ALIGNMENTS

RESULT 1	ID	Q8VXS9	PRELIMINARY	PRT	345 AA.
AC	Q8VXS9				
DT	01-MAR-2002 (T-EMBLprel. 20, Created)				
DT	01-MAR-2002 (T-EMBLprel. 20, Last sequence update)				
DT	01-OCT-2003 (T-EMBLprel. 25, Last annotation update)				
DE	Dihydroorotase.				
GN	Name=dhase;				
OS	Beta vulgaris (Sugar beet).				
OC	Bakeryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;				
OC	Caryophyllales; Amaranthaceae; Beta.				
OX	NCBI_TaxID=161934;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Kanounou R., Serrano R., Ros R.;				
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.				
CC	-1- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-				
CC	aspartate.				
CC	-1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).				
CC	-1- PATHWAY: Pyrimidine biosynthesis; third step.				
CC	-1- SIMILARITY: Belongs to the DHOase family.				
DR	EMBL; AJ251897; CAC80990.1; -.				
DR	HSSP; P05020; 1079				
DR	GO; GO:0004151; F:dihydroorotase activity; IEA.				
DR	GO; GO:0016787; F:hydrolyase activity; IEA.				
DR	GO; GO:0019856; P:pyrimidine base biosynthesis; IEA.				
DR	GO; GO:0006221; P:pyrimidine nucleotide biosynthesis; IEA.				
DR	Pfam; PF01979; Amidohydro 1; 1.				
DR	TIGRFams; TIGR00856; PYRC_dimer; 1.				
DR	PROSITE; PS00482; DIHYDROOROTASE_1; 1.				
DR	PROSITE; PS00483; DIHYDROOROTASE_2; 1.				
KW	Hydrolyase; Metal-binding; Pyrimidine biosynthesis; zinc.				
SQ	SEQUENCE 345 AA; 38297 MW; 8CA937E20895763F CRC64;				
Query Match					
Best Local Similarity 84.3%; Score 1528; DB 2; Length 345;					
Matches 282; Conservative 34; Mismatches 28; Indels 0; Gaps 0;					
Qy	1	MELSTITQDDMHLLHRDGVLLKAVSHSAHFGRAIVMPNLPKPIITTTAAVAAYREALK 60			
Db	1	MELTUTRPDDMHLLHRDGLAAVAAPSAHFGRAIVMPNLPKPIITTTGAIAARKIME 60			
Qy	61	SLPVDSPFNPMLTYLTDTPSPMEIKLARSQVFGVKLYPAGATTNSQDGVTLFGKCL 120			
Db	61	VLPPDSDFPNPLMTLYLTDTPSPMEIKLARSQVAVKLYPAGATTNSQDGVTLFGKCL 120			
Qy	121	PVLQEMFENHNPFLVHGEVTPNPEVMDREKVFETETLRLPYQKFPOLKVMHEHTTIDA 180			
Db	121	PVLEEMAEQNPFLVHGEVTPDVPDIDREKVFETESVLRPLQLKPOLKVMHEHTTIDA 180			
Qy	181	VKFVESCTEGFVAATVPQHLVLRNSLFOGGLQPHNYCPLVLRKEIHRALVSAVTSGS 240			

DB 181 VKFIESGNGNVAATVTPQHVLVNRNSLFOGGLQPHNYCLPULKEITHQALVSAVTSGS 240
 QY 241 KRFLGTDSAPHRRRRECCGCGAGIYNAPVALSVYAKVFEKEMALDKLEAFSTFNGPDF 300
 DB 241 KQFLGTDSAPHRRRRECCGCGAGIYNAPVALSVYAKVFEKEMALDKLEAFSTFNGPDF 300
 QY 301 YGLPRNNSKITLSTKTPMKVPEFSFYASGDIIPMFACEMLDWLP 344
 DB 301 YGLPRNNSKITLSTKTPMKVPEFSFYASGDIIPMFACEMLDWLP 344
 RESULT 2
 ID PYRC_ARATH STANDARD; PRT; 377 AA.
 AC 004904;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 25-OCT-2004 (Rel. 45, Last annotation update)
 GN Name:PYR4; OrderedLocustNames=At4g22930; ORFNames=F7H19.110;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. Columbia;
 RA Zhou L., Lacroute F., Thornbury R.W.;
 RT "Characterization of the Arabidopsis thaliana cDNA encoding
 RL (ex) Plant Gene Register PR97-115.
 [2]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. Columbia;
 RA MEDLINE=20083488; PubMed=10617198; DOI=10.1038/47134;
 RA Mayer K.F.X., Schaefer C., Wambutt R., Murphy G., Volckaert G.,
 RA Polt T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
 RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Teryn N.,
 RA Weichselgartner M., de Simone V., Obermaier R., Maché R., Mueller M.,
 RA Kreis M., Delaney M., Zimmermann W., Wedler H., Ridley P.,
 RA Reichert B., Porteleite D., Perez-Alonso M., Bouvier M., Bancroft I.,
 RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham J., Robben J.,
 RA Van der Schueren J., Gymnopoulos B., Chuang Y.-J., Vandenbusche F.,
 RA Breken M., Weljens I., Voet M., Baetens I., Aert R., Defoor E.,
 RA Weizsaecker T., Bothe G., Kampsperger U., Hilbert H., Braun M.,
 RA Holzner E., Brandt A., Peters S., van Staveren M., Dirke W.,
 RA Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetler P.,
 RA Berner S., Hempel S., Feldpausch M., Lambertz S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin J., Quail M.A., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIlroy K., Mayes R.,
 RA Pettit A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Blocher H., Scharte M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Faltmann B., Grandjean K., Danner D., Herzl A.,
 RA Neumann S., Argitrou A., Vitale D., Ligouri R., Piravandi E.,
 RA Massenet O., Quigley F., Clabaud G., Muehlen A., Felber R.,
 RA Chhabra S., Hiller R., Schmidt W., Lechman A., Aubourg S.,
 RA Cheffor F., Cooke R., Berger C., Montfort A., Casacuberta E.,
 RA Gieronsa T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
 RA Zaccaria P., Bevan D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Parnell L., Dedha N., Gnoj L., Schütz K., Huang B., Spiegel L.,
 RA Sekhon M., Murray J., Shee P., Cordes M., Abu-Threideh J.,
 RA Stoeckling T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Lacroix P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley J., Fulton L., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
 RA Nelson J., Speith J., Ryan E., Andrews S., Geisler C., Layman D.,

RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matexo A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Haegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M.A., Marienssen R., McCombie W.R.,
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777 (1999).
 CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
 aspartate.
 CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -1- PATHWAY: Pyrimidine biosynthesis; third step.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial (potential).
 CC -1- SIMILARITY: Belongs to the Dhase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.embnet.org/licenses/embnet.html> or send an email to license@ebi.ac.uk).
 CC
 DR EMBL; AF000146; AAB71134.1; -
 DR EMBL; AL031018; CAA19808.1; -
 DR EMBL; AL161558; CAB79248.1; -
 DR PIR; T05124; T05124.
 DR HSSP; P05020; 1379.
 DR InterPro; IPR006680; Amdihydro_1.
 DR InterPro; IPR004721; Dhodimr.
 DR InterPro; IPR002195; Dihydroorotase.
 DR Pfam; PF01979; Amdihydro_1; 1.
 DR TIGRFAMs; TIGR00856; PYRC_dimer; 1.
 DR PROSITE; PS00482; DIHYDROOROTASE_1; 1.
 DR PROSITE; PS00483; DIHYDROOROTASE_2; 1.
 KW Hydrolyase; Metal-binding; Mitochondrion; Pyrimidine biosynthesis;
 KW Transist peptide; zinc.
 FT TRANSIT 1
 FT CHAIN 1 377
 FT METAL 44 44
 FT METAL 46 46
 FT METAL 130 130
 FT METAL 168 168
 FT METAL 206 206
 FT METAL 280 280
 FT METAL 280 280
 SQ SEQUENCE 377 AA; 4493 MW; 783C22BE5581DEB CRC64;
 Query Match 81.0%; Score 1467; DB 1; Length 377;
 Best Local Similarity 78.4%; Pred. No. 9e-111;
 Matches 269; Conservative 41; Mismatches 33; Indels 0; Gaps 0;
 QY 1 MEISTOPDDMLHLDGDLVLAAYVSHSNAHFGRAIWPNNKPTTTAAVAAREALIK 60
 DB 33 MEITTPQDDMLHLDGDLVLAAYVSHSNAHFGRAIWPNNKPTTTAAVAAREALIK 92
 QY 61 SLTPVSDPFLPLTLTLTPTSPMEIKLAREGVYVGVCLYAGATNNSQDVTDLFGKCL 120
 DB 93 ALPSSSPPLMTLTLTPTSPMEIKLAREGVYVGVCLYAGATNNSQDVTDLFGKCL 152
 QY 121 PVLQEMVHNHMLLVHGVSTNPEVDMFDEKVFLETVLRPLVQVFPOLKVMMEHTTIDA 180
 DB 153 PVLQEMVHNHMLLVHGVSTNPEVDMFDEKVFLETVLRPLVQVFPOLKVMMEHTTIDA 212
 QY 181 VYFVSCTEGFAAIVTTPQHVLVNRNSLFOGGLQPHNYCLPULKEITHQALVSAVTSGS 240
 DB 213 VYFVSCTEGFAAIVTTPQHVLVNRNSLFOGGLQPHNYCLPULKEITHQALVSAVTSGS 272
 QY 241 KRFLGTDSAPHRRRRECCGCGAGIYNAPVALSVYAKVFEKEMALDKLEAFSTFNGPDF 300
 DB 273 KRFLGTDSAPHRRRRECCGCGAGIYNAPVALSVYAKVFEKEMALDKLEAFSTFNGPDF 332
 QY 301 YGLPRNNSKITLSTKTPMKVPEFSFYASGDIIPMFACEMLDWLP 343

Db 333 YGLPRNSSKITLKSPKVPDVNFPPGELVPMFAGETLQMDP 375

RESULT 3

094J38 PRELIMINARY; PRT; 359 AA.

AC 094J38; 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, last annotation update)
DE Putative dihydroorotase.
GN Name=P0481E12.10;
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OC NCBI_TaxID=3947;
OX [1]
RP SEQUENCE FROM N.A.
RX PubMed=12447438; DOI=10.1038/nature01184;
RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
RA Wu J., Nimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,
RA Hijiwhita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
RA Ikono M., Itoh S., Itoh T., Itoh Y., Iwabuchi A., Kamlya K.,
RA Karagawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,
RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
RA Shimokawa T., Shomura A., Song J., Takazaki Y., Teresawa K., Tsuji K.,
RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yukawa K.,
RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
RA Yano M., Jiang J., Gojobori T.;
RT "The genome sequence and structure of rice chromosome 1.";
RL Nature 420:312-316(2002).
CC -1- CATALYTIC ACTIVITY. (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC -1- COFACTOR. Binds 2 zinc ions per subunit (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; third step.
CC -1- SIMILARITY: Belongs to the DHOase family.
DR EMBL; AP003076; BAB56025.1; -.
DR HSSP; P05020; 1J79.
DR Gramene; Q94J38; -.
DR GO; GO:0004151; F:dihydroorotase activity; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0019856; P:pyrimidine base biosynthesis; IEA.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR004721; DHodimr.
DR InterPro; IPR002195; Dihydroorotase.
DR Pfam; PF01979; Amidohydro_1; 1.
DR TIGRFAMs; TIGR00856; PYRC_dimer; 1.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKOWN_1.
DR PROSITE; PS00483; DIHYDROOROTASE_2; 1.
KW Hydrolyase; Metal-binding; Pyrimidine biosynthesis; Zinc.
SQ SEQUENCE 359 AA; 39734 MW; AD331AF3775F863 CRC64;

Query Match 80.5%; Score 1458; DB 2; Length 359;
Best Local Similarity 78.1%; Pred. No. 4.6e-110;
Matches 268; Conservative 39; Mismatches 36; Indels 0; Gaps 0;

QY 2 ELSTQPDMDHMLRDGDVLRKAVVSHSHHFGRAIVMNLKRPITTTAAVAAYEALIKS 61
DB 12 ELITRPDMDHMLREGVLAALVLRPHSAMHFGRAIVMNLKRPVTTTBALAEYEEILRA 71
QY 62 LPVDSDFNPLMTLYLTDTTSPMEIKLAESQVDFGVKLYPAGATTNSODGVTLDFGKLP 121
DB 72 LPPGSNFFVPLMTLYLTDTTSPMEIKLAESQVDFGVKLYPAGATTNSODGVTLDFGKLP 131
QY 122 VLQGVNHNMPPLVHGEVTVNPEVDMFDEKVFIEFVLRPLVOKFPOLKVVMHVTITD 181
DB 132 VLEEMARQEMPLVHGEVTVNPEVDMFDEKVFIEFVLRPLVOKFPOLKVVMHVTITD 191

QY 182 KEVESCTEGFVAATVTPHLLVLRNSLFGGGLQPHNYCLPVLRKREIHRREALVSAVTSGRK 241
DB 192 NFVESCKSGHVAAVTPPHLLVLRNSLFGGGLQPHNYCLPVLRKREIHRREALVSAVTSGRK 251

QY 242 REFLGTSAPHDRRRKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFTSFNGPDY 301
DB 252 QYFLGTSAPHDRRRKSCGCGAGIYNAPVALSVYAKVEKENALDKLEAFTSFNGPDY 311

QY 302 GLPRNSSKITLKSPKVPDVNFPPGELVPMFAGETLQMDP 344
DB 312 GLPRNSSKITLKSPKVPDVNFPPGELVPMFAGETLQMDP 354

RESULT 4

082WF3 PRELIMINARY; PRT; 349 AA.

AC 082WF3; 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, last annotation update)
DE Dihydroorotase homodimeric type (EC 3.5.2.3).
GN Name=pyrC; OrderedlocusNames=NE0727;
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OC NCBI_TaxID=915;
OX [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 19718 / IFO 14298;
RX MEDLINE=22586410; PubMed=12700255;
DOI=10.1128/JB.185.9.2759-2773.2003;
RA Chain P., Lamedin J.B., Larimer F.W., Regala W., Lao V., Land M.L.,
RA Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayaveda-Soto L.A.,
RA Arciero D.M., Holmes N.B., Whitaker M.M., App D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
RT obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773(2003).
CC EMBL; BX321858; CAD84638.1; -.
DR HSSP; P05020; 1J79.
DR GO; GO:0004151; F:dihydroorotase activity; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0019856; P:pyrimidine base biosynthesis; IEA.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR004721; DHodimr.
DR InterPro; IPR002195; Dihydroorotase.
DR Pfam; PF01979; Amidohydro_1; 1.
DR TIGRFAMs; TIGR00856; PYRC_dimer; 1.
DR PROSITE; PS00482; DIHYDROOROTASE_1; 1.
DR PROSITE; PS00483; DIHYDROOROTASE_2; 1.
KW Complete proteome; Hydrolyase.
SQ SEQUENCE 349 AA; 39083 MW; C2603ECB255D6703 CRC64;

Query Match 59.8%; Score 1083; DB 2; Length 349;
Best Local Similarity 59.9%; Pred. No. 1.3e-79;
Matches 205; Conservative 54; Mismatches 79; Indels 4; Gaps 3;

QY 2 ELSTQPDMDHMLRDGDVLRKAVVSHSHHFGRAIVMNLKRPITTTAAVAAYEALIKS 61
DB 7 KLTFTRRDDHMLRDGNARSVLRPTARFARAIIMNLKRPVITTEQAAAYEVRILSA 66
QY 62 LPVD-SDFNPLMTLYLTDTTSPMEIKLAESQVDFGVKLYPAGATTNSODGVTLDFGKC 119
DB 67 LPAELAGRFBPLMTLYLTDTTSPMEIKLAESQVDFGVKLYPAGATTNSODGVTLDFGKC 125
QY 120 LPLVQGVNHNMPPLVHGEVTVNPEVDMFDEKVFIEFVLRPLVOKFPOLKVVMHVTITD 179
DB 126 EATLEKREIDMPLVHGEVTVNPEVDMFDEKVFIEFVLRPLVOKFPOLKVVMHVTITD 185
QY 180 AKVEESCTEGFVAATVTPHLLVLRNSLFGGGLQPHNYCLPVLRKREIHRREALVSAVTS 239
DB 186 AVEFVQTAAPNR-IAATITTAHMLNRNALFTGGLRPHHYCLPVLRKREIHRREALVSAVTS 244

OY 240 SKREFLGTDSAPHDRRRECCGCGAGIYNAPVALSVYAKVEKENALDLLEAFSTNGSD 299
 Db 245 HSREFLGTDSAPHDRRRECCGCGAGIYNAPVALSVYAKVEKENALDLLEAFSTNGSD 304
 OY 300 FYGLPRNRSKIKLSKTPKVPKVPESFSYASGDIIPMFAGEMLDW 341
 Db 305 FYGLPRNRSKIKLSKTPKVPKVPESFSYASGDIIPMFAGEMLDW 346

RESULT 5

ID PYRC RALSO STANDARD; PRT; 344 AA.
 AC OBY249;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Dihydroorotase (EC 3.5.2.3) (DHORase)
 GN Name-pyrc; Ordered locus names=RS00487; ORFNames=RS05822;
 OS Ralstonia solanacearum (Pseudomonas solanacearum)
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 OX NCBI_TaxID=305;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM1000;
 RX MEDLINE=21681879; PubMed=11823852; DOI=10.1038/415497a;
 RA Salanoubat M., Genin S., Artiguenave F., Gonzy J., Mangenot S.,
 RA Arlat M., Billaud A., Broctier P., Camus J.C., Caticolico L.,
 RA Chaudier M., Choissne N., Claudel-Renard C., Cunne S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schick T.,
 RA Sigulier P., Thebault P., Whalen M., Winkler P., Levy W.,
 RA Weisenbach J., Boucher C.A.,
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 CC -1 CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
 CC aspartate.
 CC -1 COFACTOR: Binds 2 zinc ions per subunit (By similarity).
 CC -1 PATHWAY: Pyrimidine biosynthesis; third step.
 CC -1 SUBUNIT: Homodimer (By similarity).
 CC -1 SIMILARITY: Belongs to the DHORase family. Subfamily 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL646059; CAD14015.1; -
 DR HSPR; P05020; I079.
 DR HAMAP; MF_00219; -; 1.
 DR Interpro; IPR006680; Amidohydro_1.
 DR Interpro; IPR004721; DHODmr.
 DR Interpro; IPR002195; Dihydroorotase.
 DR Pfam; PF01979; Amidohydro_1; 1.
 DR TIGRFAMs; TIGR00856; Pyrc_dimer; 1.
 DR PROSITE; PS00482; DIHYDROOROTASE 2; 1.
 DR PROSITE; PS00483; DIHYDROOROTASE 2; 1.
 KW Complete proteome; Hydrolase; Metal-binding; Pyrimidine biosyntheses;
 KM Zinc.
 FT METAL 14 14 Zinc 1 (By similarity).
 FT METAL 16 16 Zinc 1 (By similarity).
 FT METAL 100 100 Zinc 1 and 2 (By similarity).
 FT METAL 137 137 Zinc 2 (By similarity).
 FT METAL 175 175 Zinc 2 (By similarity).
 FT METAL 248 248 Zinc 1 (By similarity).
 SQ SEQUENCE 344 AA; 37816 MW; 4BB7FD71912CF610 CRC64;

Query Match 58.6%; Score 1061; DB 1; Length 344;
 Best Local Similarity 60.2%; Pred. No. 8e-77;
 Matches 204; Conservative 45; Mismatches 88; Indels 2; Gaps 2;

OY 3 LSTIOPDDMHILRDGDVILKAVYSHSAHFGRAIVMPKIPITTTAAVAAYREALIKSL 62
 Db 5 LITVRPDDMHILRDGDVILKAVYSHSAHFGRAIVMPKIPITTTAAVAAYREALIKSL 64
 OY 63 PVDSDPNPLMTLYLTDTTSPMEIKLARSQVGVKLPAGATNNSOGVYTLDFGKCLPV 122
 Db 65 PAGRFEPLMTLYLTDTTSPMEIKLARSQVGVKLPAGATNNSOGVYTLDFGKCLPV 123
 OY 123 LOEMVEHNPPLVHGVTNPEVDMFDRKVFIEYLRPLVQKFPOLKVMHEHTTIDAVK 182
 Db 124 LEMQDVGNPLVHGVTNPEVDMFDRKVFIEYLRPLVQKFPOLKVMHEHTTIDAVK 183
 OY 183 FVESCTEGFVAATVTPHOLVILNRNLSFGQGLQPHNYCLPVKRETHREALVSATSSGR 242
 Db 184 YVRD-AQGVGATITFAHHLVILNRNLSFGQGLQPHNYCLPVKRETHREALVSATSSGR 242
 OY 243 FFLGTDSPAHQDRRRECCGCGAGIYNAPVALSVYAKVEKENALDLLEAFSTNGSD 302
 Db 243 FFLGTDSPAHQDRRRECCGCGAGIYNAPVALSVYAKVEKENALDLLEAFSTNGSD 302
 OY 303 LPRNRSKIKLSKTPKVPKVPESFSYASGDIIPMFAGEMLDW 341
 Db 303 LPRNRSKIKLSKTPKVPKVPESFSYASGDIIPMFAGEMLDW 341

RESULT 6

ID Q62H45 PRELIMINARY; PRT; 352 AA.
 AC Q62H45;
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DE Dihydroorotase, homodimeric type (EC 3.5.2.3).
 GN Name-pyrc; ORFNames=BMA2422;
 OS Burkholderia mallei ATCC 23344.
 OC Burkholderia mallei ATCC 23344.
 OC Burkholderiaceae; Burkholderiales;
 OC Burkholderia mallei ATCC 23344.
 OX NCBI_TaxID=243160;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 23344;
 RA Nierman W.C., Deshaizer D., Kim H.S., Tetteelin H., Nelson K.E.,
 RA Feldblyum T., Ulrich R.L., Roming C.M., Brinkc L.M., Daugherty S.C.,
 RA Davidson T.D., Deboy R.T., Dimitrov G., Dodson R.J., Durkin A.S.,
 RA Gwynn M.L., Haft D.H., Khouiri H., Kolonay J.F., Madupu R.,
 RA Mohammed Y., Nelson W.C., Radune D., Romero C.M., Sarrica S.,
 RA Selengut J., Shambin C., Sullivan S.A., White O., Yu Y., Zafar N.,
 RA Zhou L., Fraser C.M.;
 RT "Structural flexibility in the Burkholderia mallei genome";
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14247-14251(2004).
 DR EMBL; CP000010; AAU49749.1; -
 KW Hydrolase.

SQ SEQUENCE 352 AA; 38177 MW; 88B2B9B04EB1F241 CRC64;
 Query Match 58.4%; Score 1058; DB 2; Length 352;
 Best Local Similarity 58.4%; Pred. No. 1.5e-77;
 Matches 199; Conservative 45; Mismatches 95; Indels 2; Gaps 1;

OY 3 LSTIOPDDMHILRDGDVILKAVYSHSAHFGRAIVMPKIPITTTAAVAAYREALIKSL 62
 Db 9 LITVRPDDMHILRDGDVILKAVYSHSAHFGRAIVMPKIPITTTAAVAAYREALIKSL 64
 OY 63 PVDSDPNPLMTLYLTDTTSPMEIKLARSQVGVKLPAGATNNSOGVYTLDFGKCLPV 122
 Db 65 PAGRFEPLMTLYLTDTTSPMEIKLARSQVGVKLPAGATNNSOGVYTLDFGKCLPV 128
 OY 123 LOEMVEHNPPLVHGVTNPEVDMFDRKVFIEYLRPLVQKFPOLKVMHEHTTIDAVK 182
 Db 124 LEMQDVGNPLVHGVTNPEVDMFDRKVFIEYLRPLVQKFPOLKVMHEHTTIDAVK 188
 OY 183 FVESCTEGFVAATVTPHOLVILNRNLSFGQGLQPHNYCLPVKRETHREALVSATSSGR 240
 Db 184 YVRD-AQGVGATITFAHHLVILNRNLSFGQGLQPHNYCLPVKRETHREALVSATSSGR 248


```
QY 241 KPEFLGTDSAPHDRRRKCCSCGAGIYNAPVALSVYAKVPEKENALDKLEAFTSPNGPDF 300
DB 249 PPEFLGTDSAPHAHKGAKKAACGAGCYTALHMLETAEAFDQAGALDKLEGFASFGADF 308
QY 301 YGLPRNSKIKLSKTPMKVPESSVYASGDIIPFAEMLMD 341
DB 309 YGLPRSAETVTLRRETWELPREIDAGAPVPLRGEALGM 349

RESULT 7
ID Q63QW0 PRELIMINARY; PRT; 352 AA.
AC Q63QW0;
DT 25-OCT-2004 (Tremblrel. 28, Created)
DT 25-OCT-2004 (Tremblrel. 28, Last sequence update)
DE 25-OCT-2004 (Tremblrel. 28, Last annotation update)
DE Dihydroorotase (EC 3.5.2.3).
GN Name=pyrC; ORFNames=BPSL2914;
OS Burkholderia pseudomallei K96243.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=272560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K96243;
RX PubMed=15377794;
RA Holden M.T.G., Tibbali R.W., Peacock S.J., Cerdeno-Tarraga A.M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bason N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cheever I.,
RA Chillingworth T., Cronin A., Crossell B., Davis P., Deshaizer D.,
RA Felwell T., Fraser A., Hance Z., Hauser H., Holtroyd S., Jørgels K.,
RA Keith K.B., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabbittowitch E., Rutherford K., Sanders M., Simmonds M.,
RA Songerivai S., Stevens K., Tumapa S., Veerachaveest M.,
RA Whitehead S., Yeats C., Bartell B.G., Oyston P.C.F., Parkhill J.,
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
KW EMBL, BX571965; CAH36924.1; -.
SQ SEQUENCE 352 AA; 38177 MW; 88B2B9B04EB1F241 CRC64;

Query Match 58.4%; Score 1058; DB 2; Length 352;
Best Local Similarity 58.4%; Pred. No. 1.5e-77;
Matches 199; Conservative 45; Mismatches 95; Indels 2; Gaps 1;

QY 3 LSTITPPDDMHLHLDGDLKAVVSHSAHFGRAIVMPLKPPITTTAAAVYREALIKSL 62
DB 9 LTLARPPDDMHLHLDGDLKAVVSHSAHFGRAIVMPLKPPITTTAAAVYREALIKSL 68
QY 63 PVSDDFRPLMTLYLTDTTSPMEIKLARESOVVGKLYPAGATTNSQDGVTLDFGKCLPV 122
DB 69 PAGMTFEPPLMTLYLTDTNPADRIARRSGCVHGKLYPAGATTNSDAGVTLDFGKCAKT 128
QY 123 LOEWEHNMPLLVHGEVTNPEVDMFDRKVFIEFTVLRPLVOKFPOLKVMHEHTTIDAVK 182
DB 129 LEMOEWGMPLLVHGEVTNPEVDMFDRKVFIEFTVLRPLVOKFPOLKVMHEHTTIDAVK 188
QY 183 FVE--SCTEGFVAATVTPQHLVLNRNSLFGGLOPHNYCFVLKREIHEALVSAVTSGS 240
DB 189 YVRDADAASGRIGATITAHHLVYRNAMFEGGIRPHYVCLVPLKRETHIRILVSAATSGN 248
QY 241 KRFLGTDSAPHDRRRKCCSCGAGIYNAPVALSVYAKVPEKENALDKLEAFTSPNGPDF 300
DB 249 PRFLGTDSAPHAHKGAKKAACGAGCYTALHMLETAEAFDQAGALDKLEGFASFGADF 308
QY 301 YGLPRNSKIKLSKTPMKVPESSVYASGDIIPFAEMLMD 341
DB 309 YGLPRSAETVTLRRETWELPREIDAGAPVPLRGEALGM 349
```

```
RESULT 8
PYRC_SYNY3 STANDARD; PRT; 342 AA.
ID PYRC_SYNY3
AC P74438;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Dihydroorotase (EC 3.5.2.3) (DHOase).
GN Name=pyrC; OrderedLocustNames=slr0406;
OS Synecocystis sp. (Strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97061201; PubMed=8905231;
RX Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Saenamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nartuo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; third step.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: Belongs to the DHOase family. Subfamily 1.

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL, D90915; BAA18539.1; -.
DR PIR, S76410; S76410.
DR HSSP, P05020; 1J79.
DR HAMAP, MF_00219; -; 1.
DR InterPro, IPR006680; Amidohydro_1.
DR InterPro, IPR004721; DHOdimr.
DR InterPro, IPR002195; Dihydroorotase.
DR Pfam, PF01979; Amidohydro_1; 1.
DR TIGRFAMs, TIGR00856; pyrC_dimer; 1.
DR PROSITE, PS00482; DIHYDROOROTASE_1; 1.
DR PROSITE, PS00483; DIHYDROOROTASE_2; 1.
DR Complete proteome; Hydrolyase; Metal-binding; Pyrimidine biosynthesis;
KW Zinc.
FT METAL 13 13 Zinc 1 (By similarity).
FT METAL 15 15 Zinc 1 (By similarity).
FT METAL 99 99 Zinc 1 and 2 (By similarity).
FT METAL 136 136 Zinc 2 (By similarity).
FT METAL 174 174 Zinc 2 (By similarity).
FT METAL 246 246 Zinc 1 (By similarity).
SQ SEQUENCE 342 AA; 38127 MW; 88A572814296B121 CRC64;

Query Match 58.3%; Score 1056.5; DB 1; Length 342;
Best Local Similarity 58.8%; Pred. No. 1.9e-77;
Matches 200; Conservative 52; Mismatches 85; Indels 3; Gaps 2;

QY 2 ELSTTPDDMHLHLDGDLKAVVSHSAHFGRAIVMPLKPPITTTAAAVYREALIKSL 61
DB 3 KLTITRPPDDMHLHLDGDLKAVVSHSAHFGRAIVMPLKPPITTTAAAVYREALIKSL 62
QY 62 LPVSDDFRPLMTLYLTDTTSPMEIKLARESOVVGKLYPAGATTNSQDGVTLDFGKCLPV 121
DB 63 IPAGGGEPEPLMTLYLTDTNPDEEIIAAKASQFVAAYVYYPAGATTNSDFGVTIDH-RCDA 121
QY 122 VLQEMVEHNMPLLVHGEVTNPEVDMFDRKVFIEFTVLRPLVOKFPOLKVMHEHTTIDAV 181
```

```

Db 122 VLAAMEQVDLP.LLHGEVTSNDIPFREKVFLEKILIPREKFPPLRVLEHITTSADV 181
Qy 182 KEVSSCTEGFVAATVTOHVLVNRSLFQCGLOPHNYCPLVREIHRALVSATVSGSK 241
Db 182 QFVLSANN--IAATITFOHLLFRNMLFKGIGLPHFYCLPLIKREBRLSLHAATSGNP 239
Qy 242 RFLGTDAPBHDRRKESCGCAGIYNAPVALSVKAVKEXNALDKLEAFTSFNGDPFY 301
Db 240 KFLFGTDSAPBHRNSKESLGGCAGCYSLHAMELYLEAFESVNSLDKLEAFASFGDPFY 299
Qy 302 GLPRNNSKITLSTKTPKVPESFSYASGDIIPMFAGEMLDW 341
Db 300 QLEPNTAQITLTMKPMRIPALFPPESGLVPLRAGEBITW 339

```

RESULT 9

```

ID 07NO12 PRELIMINARY; PRT; 361 AA.
AC 07NO12;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DE Dihydroorotase (EC 3.5.2.3).
GN Name=pyrC, OrderedLocusName=CV4330;
OS Chromobacterium violaceum.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Chromobacterium.
NCBI_TaxId=536;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12472 / DSM 30191;
RX MEDLINE=2282880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Gultarases C.T.,
RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
RA Alves-Gomes J.A., Andrade E.M., Araxipe J., de Araujo M.F.F.,
RA Astolfi-Filho S., Azevedo V., Baptista A.J., Batous L.A.M.,
RA Batista J.S., Beijo A., van den Berg C., Bogo M., Bonatto S.,
RA Borzignon J., Brigido M.M., Brito C.A., Brocchi M., Butty H.A.,
RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
RA Carvalho C.M.B., Cascado J.C.M., Cavada B.S., Chueire L.M.O.,
RA Crezzymski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
RA Fantiuch F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
RA Gazezelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
RA Gracitapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,
RA Lecl L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
RA Madeira H.M.F., Mantio G.P., Maranhao A.O., Martins W.S.,
RA di Mauro S.M.Z., de Medeiros S.R.B., Meisner R.V., Moreira M.A.M.,
RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seunarez H.N.,
RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
RA Vettore A., Wassem R., Zaha A., Simpson A.J.G.;
RT "The complete genome sequence of Chromobacterium violaceum reveals
RT remarkable and exploitable bacterial adaptability."
RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; third step.
CC EMBL: A5016925; AA061989.1; -.
CC HSSP: P05020; 1J79.
DR GO: GO:0004151, F:dihydroorotase activity; IEA.
DR GO: GO:0016787, F:pyrimidine base biosynthesis; IEA.
DR GO: GO:0018566, F:pyrimidine base biosynthesis; IEA.
DR GO: GO:0006221, P:pyrimidine nucleotide biosynthesis; IEA.
DR InterPro: IPR006680; Amidohydro_1.

```

```

DR InterPro: IPR004721; DHODImr.
DR InterPro: IPR002195; Dihydroorotase.
DR Pfam: P01979; Amidohydro_1; 1.
DR TIGRFAMs: TIGR00856; pyrC_dimer; 1.
DR PROSITE: PS00483; DIHYDROOROTASE 2; 1.
DR Complete proteome; Hydrolyase; Metal-binding; Pyrimidine biosynthesis;
DR Zinc.
SQ SEQUENCE 361 AA; 38985 MW; 6B62FD360883FEF3 CRC64;

```

```

Query Match 58.1%; Score 1053; DB 2; Length 361;
Best Local Similarity 59.0%; Pred. No. 3.8e-77;
Matches 200; Conservative 46; Mismatches 91; Indels 2; Gaps 2;

```

```

Qy 3 LSITQPDWHLRLRGDVLKAVSAAHFGAIVMPKLPITTTAAVAAYREAILKSL 62
Db 21 LTLIRPDDWHLRLRDALAAVLPTSCOMGRALVMPKLPVTVAATAAAYRRRIILAR 80
Qy 63 PVDSDPPLMTLYLTDTSPEIKLARESOVVFVKLYPAGATTNSODGYTDLFGKCLPV 122
Db 81 PAGSAPEPLMTLYLTDKTSADEIRKAKACGFVHGKLYPACATTNSDHYGTDI-AHAMPA 139
Qy 123 LQEMVHNMPILVHGEVTPNPEVDMPREKVFLEKILIPREKFPPLRVLEHITTSADV 182
Db 140 LEAMAEAGMPILVHGEVTDADIDVDFDEAVFIEKMPKLPKLPOLKVFHEHITTTREAAE 199
Qy 183 FPESTCEGFVAATVTOHVLVNRSLFQCGLOPHNYCPLVLEIHRALVSATVSGSKR 242
Db 200 FVAAPAN-VAAITTAHLMLNRRNALFVGGRPHHYCLPVLKELHRLQALVKAATSGSAK 258
Qy 243 FFLGTDAPBHDRRKESCGCAGIYNAPVALSVKAVKEXNALDKLEAFTSFNGDPFY 302
Db 259 FFLGTDAPBHRNSKESLGGCAGCYSLHAMELYLEAFESVNSLDKLEAFASFGDPFY 318
Qy 303 LPRNNSKITLSTKTPKVPESFSYASGDIIPMFAGEMLDW 341
Db 319 LPANGRIELVKSMTVPALFPPESGLVPLRAGEBITW 357

```

RESULT 10

```

ID PYRC_VIBCH STANDARD; PRT; 342 AA.
AC 09KJ24;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 25-OCT-2004 (Rel. 45, Last annotation update)
DE Dihydroorotase (EC 3.5.2.3) (DHODase).
GN Name=pyrC, OrderedLocusName=VCA0925;
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
NCBI_TaxId=666;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=B1 Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;
RA Heidelberg J.F., Bisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gali S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.L.,
RA Emolaeva M.D., Vamathevan J.J., Bae S., Qin H., Dragoi I.,
RA Sellers P., McDonald L.A., Utecherback T.R., Fleischmann R.D.,
RA Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R.,
RA Mekalanos J.J., Venter J.C., Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC -1- COFACTOR: Binds 2 zinc ions per subunit (By similarity).
CC -1- PATHWAY: Pyrimidine biosynthesis; third step.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SIMILARITY: Belongs to the DHODase family. Subfamily 1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AB004420; AAF96822.1; ALT_INIT.
DR HSSP; P05020; 1d79.
DR TIGR; VCA0925; -.
DR HAMAP; MF_00219; -; 1.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR004721; DHODIm.
DR InterPro; IPR002195; Dihydroorotase.
DR Pfam; PF01979; Amidohydro_1; 1.
DR TIGRfam; TIGR00856; pyrC_dimer; 1.
DR PROSITE; PS00482; DIHYDROOROTASE_1; 1.
DR PROSITE; PS00483; DIHYDROOROTASE_2; 1.
KW Complete proteome; Hydrolase; Metal-binding; Pyrimidine biosynthesis;
KW Zinc.
FT METAL 13 13 Zinc 1 (By similarity).
FT METAL 15 15 Zinc 1 (By similarity).
FT METAL 98 98 Zinc 1 and 2 (By similarity).
FT METAL 135 135 Zinc 2 (By similarity).
FT METAL 173 173 Zinc 2 (By similarity).
FT METAL 246 246 Zinc 1 (By similarity).
SQ SEQUENCE 342 AA; 37461 MW; 9PAD2DE6E1818DA CRC64;

Query Match 57.4%; Score 1039.5; DB 1; Length 342;
Best Local Similarity 59.0%; Pred. No. 4.5e-76;
Matches 200; Conservative 45; Mismatches 91; Indels 3; Gaps 3;

QY 3 LSTTPDDMHLHLDGDLVKAIVSHSAHFGRAIWMENLKRPITTTAAVAAYREAILKSL 62
DB 4 LITRPDDMHLHLDGDLVKAIVSHSAHFGRAIWMENLKRPITTTAAVAAYREAILKSL 63
QY 63 PVSDENPLMTLYLTDTTSPMEIKLARESOVGVKLYPRGATNNSODGVTDLFGKCLPV 122
DB 64 P-QAHFEPMLALYLTDNTPSEIRKAKSGKVAAKLYPAGATNNSOGVTSI-KNITVPV 121
QY 123 LQEMVENMPLVHGEVTNPEVDMFDEKVFIEFTVLRPLVOKFPOLKVMEHVTITDAVK 182
DB 122 LQAMQEGMQLLHGEVTNPEVDMFDEKVFIEFTVLRPLVOKFPOLKVMEHVTITDAVK 181
QY 183 FVESCTEGFAVATVTPQHLVLNRSLSFGGLQPHNYCLPVLRKEIHREALVSAVTSQSKR 242
DB 182 FVQAGDN-VAGITTAHLLFNRRHMLVGIIRPHFYCLPIKRAITHQALVAATSGSKK 240
QY 243 FFLGTDSAPHDRRRKESCGCAGIYNAPVALSVYAKYFEKNALDKLEATSPNGDPFG 302
DB 241 FFLGTDSAPHAQGRKESACGAGSYTAHAALLEYAEVFEAKLENTAEAFSPNGDPFG 300
QY 303 LPRNNSKIKLSKTPKPKPESFSYASGDIIMPAGEMLDM 341
DB 301 LPRNQTIVTLTKQAMPVAESMPFGSDIVVPIRAGENIEM 339

RESULT 11
Q6LP17 PRELIMINARY; PRT; 342 AA.
AC Q6LP17;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Putative dihydroorotase.
GN Name=C44330; OrderedlocusNames=PBRAA405.
OS Photobacterium profundum (Photobacterium sp. (strain SS9)).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Photobacterium.
OX NCBI_TaxID=74109;
RN [1]
RP SEQUENCE FROM N.A.
RA Vezzi A., Campanaro S., D'Angelo M., Simonato F., Vitulo N., Lauro F.,

RA Cestaro A., Malacrida G., Simonati B., Cannata N., Bartlett D.,
RA Valle G.;
RT "Genome analysis of Photobacterium profundum reveals the complexity of
RT high pressure adaptations.";
RT Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL; CR378670; CAG20789.1; -.
DR GO; GO:0004151; P.dihydroorotase activity; IEA.
DR GO; GO:0016787; P.dihydroorotase activity; IEA.
DR GO; GO:0019856; P.pyrimidine base biosynthesis; IEA.
DR InterPro; IPR006680; Amidohydro_1.
DR InterPro; IPR004721; DHODIm.
DR InterPro; IPR002195; Dihydroorotase.
DR Pfam; PF01979; Amidohydro_1; 1.
DR TIGRfam; TIGR00856; pyrC_dimer; 1.
DR PROSITE; PS00482; DIHYDROOROTASE_1; 1.
DR PROSITE; PS00483; DIHYDROOROTASE_2; 1.
KW Complete proteome.
SQ SEQUENCE 342 AA; 37603 MW; AFE28ABEFEBDA022 CRC64;

Query Match 55.9%; Score 1012.5; DB 2; Length 342;
Best Local Similarity 56.0%; Pred. No. 7e-74;
Matches 190; Conservative 57; Mismatches 89; Indels 3; Gaps 3;

QY 3 LSTTPDDMHLHLDGDLVKAIVSHSAHFGRAIWMENLKRPITTTAAVAAYREAILKSL 62
DB 4 LITRPDDMHLHLDGDLVKAIVSHSAHFGRAIWMENLKRPITTTAAVAAYREAILKSL 63
QY 63 PVSDENPLMTLYLTDTTSPMEIKLARESOVGVKLYPRGATNNSODGVTDLFGKCLPV 122
DB 64 P-QGNSPLMVLVLTNTSABEIKKASGHVAAKLYPAGATNNSOGVTSI-DHITRPA 121
QY 123 LQEMVENMPLVHGEVTNPEVDMFDEKVFIEFTVLRPLVOKFPOLKVMEHVTITDAVK 182
DB 122 LQAMQEGMQLLHGEVTNPEVDMFDEKVFIEFTVLRPLVOKFPOLKVMEHVTITDAVK 181
QY 183 FVESCTEGFAVATVTPQHLVLNRSLSFGGLQPHNYCLPVLRKEIHREALVSAVTSQSKR 242
DB 182 FVNNAGN-VGAIITTAHLLFNRRHMLVGIIRPHFYCLPIKNIHQDALVKAATSGNPK 240
QY 243 FFLGTDSAPHDRRRKESCGCAGIYNAPVALSVYAKYFEKNALDKLEATSPNGDPFG 302
DB 241 FFLGTDSAPHAQGRKESACGAGSYTAHAALLEYAEVFEADALDKLEATSPNGDPFG 300
QY 303 LPRNNSKIKLSKTPKPKPESFSYASGDIIMPAGEMLDM 341
DB 301 LPRNTDTITLKGSWNPETMAFGDEVVPIRAGENIEM 339

RESULT 12
Q7M4A4 PRELIMINARY; PRT; 359 AA.
AC Q7M4A4;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Dihydroorotase (EC 3.5.2.3).
GN Name=pyrC; OrderedlocusNames=BP3762;
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=12822 / ATCC BAA-587;
RC MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RX Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Achtman M., Atkin R., Baker S., Basham D., Bacon N., Chevreton I.,
RA Chillingworth T., Collins M., Cronin H., Davis P., Doggett J.,
RA Fellwell T., Goble A., Hamlin M., Hauser H., Holtroyd S., Jørgen K.,
RA Leather S., Moule S., Norbertczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,

RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
 RT "Comparative analysis of the genome sequences of *Bordetella pertussis*,
 RT *Bordetella parapertussis* and *Bordetella bronchiseptica*.",
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL: BX640434; CAE39045.1; -.
 DR HSSP: P05020; 1J79.
 DR GO: GO:0004151; F: dihydrocrotonase activity; IEA.
 DR GO: GO:0016787; F: hydrolase activity; IEA.
 DR GO: GO:0019856; P: pyrimidine base biosynthesis; IEA.
 DR InterPro: IPR006580; Amidohydro_1.
 DR InterPro: IPR004721; Dihydroim.
 DR InterPro: IPR002195; Dihydrocrotonase.
 DR Pfam: PF01979; Amidohydro_1; 1.
 DR TIGRFAMs: TIGR00856; pyrc_dimer; 1.
 DR PROSITE: PS00482; DIHYDROCRONASE_1; 1.
 DR PROSITE: PS00483; DIHYDROCRONASE_2; 1.
 KW Complete proteome; Hydrolase.
 SQ SEQUENCE 359 AA; 38969 MW; A75138CE0C7DE97 CRC64;

Query Match 55.9%; Score 1012.5; DB 2; Length 359;
 Best Local Similarity 57.3%; Pred. No. 7.4e-74;
 Matches 197; Conservative 46; Mismatches 96; Indels 5; Gaps 3;

QY 2 ELSTQPDMDHLRDGDLKAVVSHSAHFGRAIWMNPKPITTTAAVAYR---EAI 58
 DB ELSTTRPDMDHLRDSALAEAVLDTRQFARAIVAPNRPPTTTEQALAYARIEA 72
 QY 59 LKSLPVD-SDFNPMLTYLTDTTSPMEIKLARESOVFGVLYPAGATTNSQDGVTLDFG 117
 DB LKAGGDTAAFTPLMTLYLTDTTSPAEIIVRAHESGOVAAYKLPAGATTNSDGVTLDFG 132
 QY 73 LKAGGDTAAFTPLMTLYLTDTTSPAEIIVRAHESGOVAAYKLPAGATTNSDGVTLDFG 132
 DB 118 KCLPVLQEMVHNHNPVLVHGEVTPNPEVDMFPEKVFETVLRPLVQKFPOLKVMHEVTT 177
 DB 133 KCGAALALERCQMPLVHGEVTPDAIDVFDEAVFIERWQPLRRAYPGLKVFHEITT 192
 QY 178 IDAVKFEVSCTEGFAATVTPQHLVLRNLSFQGLDPHNYCPLVLRKEITHREALVAVT 237
 DB 193 REGAHVVD-AEGTPAATITTPQHLVYRNALFTGCVAPHMYCPLIKREYHROALVEAAT 251
 QY 238 SGRKRFELGTD SAPDRRREKSCGAGIYNAPVALSVYKVEKENALDKLEAFTSPFG 297
 DB 252 SGRPRFLGTD SAPHARGLEKHAAGCAGCTALHAMELYATAPDAGRLDRLEGFASFG 311
 QY 298 PDFTGLPRNNSKITLSTPMKVPESFSYASGDIIPMAGEMLDW 341
 DB 312 PDFTGLPRNTGTTLTRREAYEIPAEVAFGDTTLVPLAGSGLGM 355

RESULT 13

Q7WFR2 PRELIMINARY; PRT; 359 AA.
 AC Q7WFR2;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Dihydrocrotonase (EC 3.5.2.3).
 GN Name:pyrc; Ordered locus names=BB84208;
 OS *Bordetella bronchiseptica* (Alcaligenes bronchisepticus).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; *Bordetella*.
 NC NCB1_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RB50 / ATCC BAA-568;
 RA MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
 RA Parkhill J., Sebatia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagsels K.,
 RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,

RA Rabbintowech E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
 RT "Comparative analysis of the genome sequences of *Bordetella pertussis*,
 RT *Bordetella parapertussis* and *Bordetella bronchiseptica*.",
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL: BX640449; CAE34572.1; -.
 DR HSSP: P05020; 1J79.
 DR GO: GO:0004151; F: dihydrocrotonase activity; IEA.
 DR GO: GO:0016787; F: hydrolase activity; IEA.
 DR GO: GO:0019856; P: pyrimidine base biosynthesis; IEA.
 DR InterPro: IPR006580; Amidohydro_1.
 DR InterPro: IPR004721; Dihydroim.
 DR InterPro: IPR002195; Dihydrocrotonase.
 DR Pfam: PF01979; Amidohydro_1; 1.
 DR TIGRFAMs: TIGR00856; pyrc_dimer; 1.
 DR PROSITE: PS00482; DIHYDROCRONASE_1; 1.
 DR PROSITE: PS00483; DIHYDROCRONASE_2; 1.
 KW Complete proteome.
 SQ SEQUENCE 359 AA; 38969 MW; A75138CE0C7DE97 CRC64;

Query Match 55.9%; Score 1012.5; DB 2; Length 359;
 Best Local Similarity 57.3%; Pred. No. 7.4e-74;
 Matches 197; Conservative 46; Mismatches 96; Indels 5; Gaps 3;

QY 2 ELSTQPDMDHLRDGDLKAVVSHSAHFGRAIWMNPKPITTTAAVAYR---EAI 58
 DB ELSTTRPDMDHLRDSALAEAVLDTRQFARAIVAPNRPPTTTEQALAYARIEA 72
 QY 59 LKSLPVD-SDFNPMLTYLTDTTSPMEIKLARESOVFGVLYPAGATTNSQDGVTLDFG 117
 DB LKAGGDTAAFTPLMTLYLTDTTSPAEIIVRAHESGOVAAYKLPAGATTNSDGVTLDFG 132
 QY 73 LKAGGDTAAFTPLMTLYLTDTTSPAEIIVRAHESGOVAAYKLPAGATTNSDGVTLDFG 132
 DB 118 KCLPVLQEMVHNHNPVLVHGEVTPNPEVDMFPEKVFETVLRPLVQKFPOLKVMHEVTT 177
 DB 133 KCGAALALERCQMPLVHGEVTPDAIDVFDEAVFIERWQPLRRAYPGLKVFHEITT 192
 QY 178 IDAVKFEVSCTEGFAATVTPQHLVLRNLSFQGLDPHNYCPLVLRKEITHREALVAVT 237
 DB 193 REGAHVVD-AEGTPAATITTPQHLVYRNALFTGCVAPHMYCPLIKREYHROALVEAAT 251
 QY 238 SGRKRFELGTD SAPDRRREKSCGAGIYNAPVALSVYKVEKENALDKLEAFTSPFG 297
 DB 252 SGRPRFLGTD SAPHARGLEKHAAGCAGCTALHAMELYATAPDAGRLDRLEGFASFG 311
 QY 298 PDFTGLPRNNSKITLSTPMKVPESFSYASGDIIPMAGEMLDW 341
 DB 312 PDFTGLPRNTGTTLTRREAYEIPAEVAFGDTTLVPLAGSGLGM 355

RESULT 14

Q7VUP3 PRELIMINARY; PRT; 359 AA.
 AC Q7VUP3;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Dihydrocrotonase (EC 3.5.2.3).
 GN Name:pyrc; Ordered locus names=BP3035;
 OS *Bordetella pertussis*.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; *Bordetella*.
 NC NCB1_TaxID=520;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tohama I / ATCC BAA-568 / NCTC 13251;
 RA MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
 RA Parkhill J., Sebatia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdano-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagsels K.,

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2005, 00:25:15 ; Search time 767 Seconds
(without alignments)
9809.637 Million cell updates/sec

Title: US-10-070-277-1

Perfect score: 1271
Sequence: 1 ttgcaaaatggagctctca.....tgaagattcctgataaaaa 1271

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseq1980s:*
2: geneseq1990s:*
3: geneseq2000s:*
4: geneseq2001as:*
5: geneseq2001bs:*
6: geneseq2002as:*
7: geneseq2002bs:*
8: geneseq2003as:*
9: geneseq2003bs:*
10: geneseq2003cs:*
11: geneseq2003ds:*
12: geneseq2004as:*
13: geneseq2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1271	100.0	1271	4	AAf61217
2	1271	100.0	1271	4	AAf61262
3	646	50.8	1743	6	AAf61459
4	281	22.1	1053	8	AAf61459
5	281	22.1	1071	10	AAf61459
6	277.2	21.6	1047	8	AAf61459
7	272.8	21.5	1053	10	AAf61459
8	272.8	21.5	110000	10	AAf61459
9	272.8	21.5	182624	10	AAf61459
10	267.4	21.0	1032	8	AAf61459
11	267.4	21.0	1041	9	AAf61459
12	265.2	20.9	1056	8	AAf61459
13	264.4	20.8	1092	8	AAf61459
14	264.2	20.8	110000	6	AAf61459
15	245.6	19.3	1047	8	AAf61459
16	243	19.1	1068	12	AAf61459
17	242.6	19.1	1197	11	AAf61459
18	242.6	19.1	1262	11	AAf61459
19	241	19.0	1047	8	AAf61459
20	238.2	18.7	1062	8	AAf61459

21	238.2	18.7	1062	8	ACA39060	ACA39060 Prokaryot
22	238.2	18.7	92407	4	AAf28549	AAf28549 Genomic f
23	236.2	18.6	1044	8	ACA43541	ACA43541 Prokaryot
24	236	18.6	1032	8	ACA41184	ACA41184 Prokaryot
25	236	18.6	1032	10	AB239013	AB239013 N. gonorr
26	236	18.6	1119	10	AB239014	AB239014 N. gonorr
27	232.8	18.3	30078	3	AAa81520	AAa81520 N. mening
28	232.8	18.3	110000	3	AAa81490_06	AAa81490_06 Continuation (7 of
29	232.8	18.3	110000	3	AAa81490_07	AAa81490_07 Continuation (8 of
30	232.8	18.3	34980	3	AAf21608	AAf21608 Neisseria
31	231.2	18.2	1035	8	ACA41757	ACA41757 Prokaryot
32	221.6	17.4	1044	8	ACA35674	ACA35674 Prokaryot
33	221.6	17.4	1059	11	ACA94623	ACA94623 Klebsiella
34	219.2	17.2	1041	8	ACA45873	ACA45873 Prokaryot
35	218.4	17.2	1047	8	ACA51824	ACA51824 Prokaryot
36	186.4	14.7	1052	8	ACA8698	ACA8698 Prokaryot
37	183.8	14.5	642	13	ACN46418	ACN46418 Cotton pr
38	177	13.9	381	12	ADP95135	ADP95135 Cotton ex
39	149	11.7	574	8	ACA11496	ACA11496 Prokaryot
40	138.8	10.9	574	13	ACN48140	ACN48140 Cotton pr
41	129.8	10.2	1266	11	ABD02404	ABD02404 Pseudomon
42	114.4	9.0	1008	8	ACA30213	ACA30213 Prokaryot
43	114.4	9.0	393	10	ACF66304	ACF66304 Pseudomon
44	105	8.3	407	8	ACA44862	ACA44862 Prokaryot
45	79.6	6.3	261	11	ABD02513	ABD02513 Pseudomon

ALIGNMENTS

RESULT 1	AAf61217	standard, cDNA; 1271 BP.
ID	AAf61217	
XX	AAf61217	
AC	25-MAY-2001	(first entry)
XX	Potato dihydro-oroctase cDNA.	
DE	Potato; dihydro-oroctase; plant; polysaccharide content; orotic acid;	
XX	uridine; starch synthesis; ss.	
KW	Solanum tuberosum.	
XX	Key	Location/Qualifiers
FT	CDS	9..1049
FT		/*tag= a
FT		/product= "dihydro-oroctase"
XX	W0200114569-A2.	
PN	01-MAR-2001.	
XX	12-AUG-2000; 2000MO-EP007884.	
PR	20-AUG-1999; 99DE-01039688.	
XX	(BADI) BASF AG.	
PA	Ehhardt T, Stltt Nigel M, Geigenberger PL, Loeff I, Zrenner R;	
PI	Schroeder M;	
XX	WPI; 2001-202938/20.	
DR	P-PSDB; AAB70773.	
XX	Increasing the polysaccharide, especially starch, content of plants, by	
PT	transforming with a sequence encoding dihydro-oroctase.	
XX	Claim 3; Page 28-30; 32p; German.	
PS	This invention describes a novel use of a DNA sequence (I) encoding	
CC	dihydro-oroctase (II) to prepare plants which have increased	

CC polysaccharide (PS) content. The invention also describes (1) a 1271 base
 CC pair DNA sequence (I) encoding potato (II), sequences that hybridize to
 CC it and its homologs (or their fragments); (2) method for preparing plants
 CC of increased PS content by expressing (I) in them; (3) method for
 CC transforming plants by introducing an expression cassette, containing
 CC promoter, signal sequence and (1), into a plant cell, callus tissue,
 CC complete plant or protoplast; and (4) plants with increased PS content
 CC containing the DNA of (1). Expression of (II) increases the content of
 CC orotic acid (a precursor of uridine nucleotides), resulting in an
 CC increased rate of starch synthesis. Typically, in potato tuber slices,
 CC addition of 10 mM orotate more than doubled the rate of starch synthesis.
 CC (1) is used to prepare plants (especially tomato, tobacco, potato,
 CC tapioca, manioc, rice, barley, oats, rye, wheat and maize) with increased
 CC PS, particularly starch, content
 XX
 SQ Sequence 1271 BP; 356 A; 244 C; 278 G; 393 T; 0 U; 0 Other;

Query Match 100.0%; Score 1271; DB 4; Length 1271;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCAAAAATGAGAGCTCTCAATACACAACTGATGATGGCATCTTCTCCGTATG 60
 DB 1 TTGCAAAAATGAGAGCTCTCAATACACAACTGATGATGGCATCTTCTCCGTATG 60
 QY 61 GTGAGTCTTAAAGGAGGAGTGTCTCTCAAGTGCATCACTTTGGAGGCAATATGCA 120
 DB 61 GTGAGTCTTAAAGGAGGAGTGTCTCTCAAGTGCATCACTTTGGAGGCAATATGCA 120
 QY 121 TGCCAAATTTGAAAGCTCTCTATCACTACGCTGCTGTATGCAATACCGGAGGCGA 180
 DB 121 TGCCAAATTTGAAAGCTCTCTATCACTACGCTGCTGTATGCAATACCGGAGGCGA 180
 QY 121 TGCCAAATTTGAAAGCTCTCTATCACTACGCTGCTGTATGCAATACCGGAGGCGA 180
 DB 121 TGCCAAATTTGAAAGCTCTCTATCACTACGCTGCTGTATGCAATACCGGAGGCGA 180
 QY 181 TATTGAATCTTTTACCTGTATAGTGAATTTCAACCTCTTATGACATTTATTTGACAG 240
 DB 181 TATTGAATCTTTTACCTGTATAGTGAATTTCAACCTCTTATGACATTTATTTGACAG 240
 QY 241 ATACAACAGCTCTATGGAATCAACATGAGAGAGAGGAGGCTGATTTTGGGGTGA 300
 DB 241 ATACAACAGCTCTATGGAATCAACATGAGAGAGAGGAGGCTGATTTTGGGGTGA 300
 QY 301 AGTTGACCTGCTGTGTCACGACAAATTTCTCAAGTGAAGTGAATCTTTTGGGA 360
 DB 301 AGTTGACCTGCTGTGTCACGACAAATTTCTCAAGTGAAGTGAATCTTTTGGGA 360
 QY 361 AGTTGACCTGCTGTGTCACGACAAATTTCTCAAGTGAAGTGAATCTTTTGGGA 360
 DB 361 AGTTGACCTGCTGTGTCACGACAAATTTCTCAAGTGAAGTGAATCTTTTGGGA 360
 QY 421 AGTTGACCTGCTGTGTCACGACAAATTTCTCAAGTGAAGTGAATCTTTTGGGA 480
 DB 421 AGTTGACCTGCTGTGTCACGACAAATTTCTCAAGTGAAGTGAATCTTTTGGGA 480
 QY 481 TAAAGCCGTGTCGAGAAATTTCCCAATTGAAGTGTGATGAGATGTTACACCA 540
 DB 481 TAAAGCCGTGTCGAGAAATTTCCCAATTGAAGTGTGATGAGATGTTACACCA 540
 QY 541 TTGATCTGTAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 DB 541 TTGATCTGTAAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 QY 601 CACAATCTGTTTGAACAGAAATTTCTTCCAAAGGAGGCTTACACCGCATATT 660
 DB 601 CACAATCTGTTTGAACAGAAATTTCTTCCAAAGGAGGCTTACACCGCATATT 660
 QY 661 ACTGCTTCCTCAAGTCTCAAAAGAGATCCACAGGAGGAGCTTGTGACCTGTAA 720
 DB 661 ACTGCTTCCTCAAGTCTCAAAAGAGATCCACAGGAGGAGCTTGTGACCTGTAA 720
 QY 721 GTGGAAGTAAAGATTTTCTTGGAGCTGATGATGATGATGATGATGATGATGATGAT 780
 DB 721 GTGGAAGTAAAGATTTTCTTGGAGCTGATGATGATGATGATGATGATGATGATGAT 780

QY 781 AGTGTCTTGGAGATGTCGTGATTTTAAATGACACCTGTAGCTTGTGATATAGCA 840
 DB 781 AGTGTCTTGGAGATGTCGTGATTTTAAATGACACCTGTAGCTTGTGATATAGCA 840
 QY 841 AGTGTCTTGGAGATGTCGTGATTTTAAATGACACCTGTAGCTTGTGATATAGCA 900
 DB 841 AGTGTCTTGGAGATGTCGTGATTTTAAATGACACCTGTAGCTTGTGATATAGCA 900
 QY 901 CAGATTTTAAAGGAGTCTTCTTGAAGACATCTCAAGATTAAGTGAATGAGAGCCATGCA 960
 DB 901 CAGATTTTAAAGGAGTCTTCTTGAAGACATCTCAAGATTAAGTGAATGAGAGCCATGCA 960
 QY 961 AGTGTCTTGGAGATGTCGTGATTTTAAATGACACCTGTAGCTTGTGATATAGCA 1020
 DB 961 AGTGTCTTGGAGATGTCGTGATTTTAAATGACACCTGTAGCTTGTGATATAGCA 1020
 QY 1021 TGCTGACCTGTTGCGGCTCTCTCTGAGATCAATTTGTGATCTTGTATGATAT 1080
 DB 1021 TGCTGACCTGTTGCGGCTCTCTCTGAGATCAATTTGTGATCTTGTATGATAT 1080
 QY 1081 GTGATTTCAACAAAGATTAAGTGAATGAGAGCCATGCA 1140
 DB 1081 GTGATTTCAACAAAGATTAAGTGAATGAGAGCCATGCA 1140
 QY 1141 ATTATCAGATGATTAATCTTCTTCAAGTGAATGAGAGCCATGCA 1200
 DB 1141 ATTATCAGATGATTAATCTTCTTCAAGTGAATGAGAGCCATGCA 1200
 QY 1201 CTCACCAAGTCTCTGATTTTGAAGTCTTAAATTTGTTGTTGATTTGAAGATTT 1260
 DB 1201 CTCACCAAGTCTCTGATTTTGAAGTCTTAAATTTGTTGTTGATTTGAAGATTT 1260
 QY 1261 ACTGATAAAA 1271
 DB 1261 ACTGATAAAA 1271

RESULT 2
 ID AAF61262 standard; cDNA; 1271 BP.
 XX AAF61262;
 AC AAF61262;
 DT 25-MAY-2001 (first entry)
 XX
 DE Potato dihydro-orotase cDNA.
 XX
 KM Potato; dihydro-orotase; inhibitor; herbicide; transgenic plant; ss.
 XX
 OS Solanum tuberosum.
 XX
 FH Key
 FT CDS
 FT Location/Qualifiers
 FT 9..1049
 FT /*tag= a
 FT /product= "dihydro-orotase"
 XX
 PN W0200118190-A2.
 PD 15-MAR-2001.
 PF 02-SEP-2000; 2000MO-EP008581.
 PR 07-SEP-1999; 99DE-01042742.
 XX
 PA (BAD1) BASF AG.
 XX
 PI Ehrhardt T, Lerchl J, Stitt Nigel M, Zrenner R, Schroeder M,
 DR WPI; 2001-235198/24.
 DR P-PSDB; AAB70779.
 XX
 PT New DNA encoding plant dihydroorotase, for producing herbicide-resistant
 PT plants and for screening for compounds that can inhibit dihydroorotase

PT and that be used as herbicides.

XX Claim 1, Page 29-31, 38pp; German.

XX
PS
XX
CC This invention describes a novel DNA sequence (I) containing the coding
CC region for a plant dihydro-oxore (DHO), comprising a sequence (S1) of
CC 1271 base pairs (bp), given in the specification. The invention also
CC describes (1) a DNA sequence (Ia) that hybridizes to (S1), or its parts
CC or derivatives produced by insertion, deletion or substitution, and
CC encodes a protein with DHO activity; (2) a protein (II) containing at
CC least 100 amino acids (aa) from a 346 aa sequence (S2), given in the
CC specification; (3) identifying substances (III) that inhibit activity of
CC plant DHO; (4) identifying substances (IIia) which act as herbicides by
CC inhibition of plant DHO; (5) a test system based on expression of (S1)
CC for identifying herbicidal inhibitors of DHO; and (6) inhibitors of plant
CC DHO identified by the system of (5). The products of the invention have
CC herbicide activity. (I) is used: (i) for recombinant production of DHO in
CC prokaryotic or eukaryotic cells; (ii) to produce a test system for
CC identifying compounds that inhibit DHO, potentially useful as herbicides;
CC and (iii) to produce transgenic plants that have increased resistance to
CC DHO-inactivating herbicides. Recombinantly expressed DHO provides a
CC reproducible and rapid (high throughput) screen for specific inhibitors
CC for use as herbicides, including compounds with widely differing degrees
CC of activity

XX Sequence 1271 BP; 356 A; 244 C; 278 G; 393 T; 0 U; 0 Other;

Query Match 100.0%; Score 1271; DB 4; Length 1271;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGCAAAATGAGAGCTCTCAATCAACAACTGATGATGGCATCTTCATCTCGGTATG 60
DB 1 TTGCAAAATGAGAGCTCTCAATCAACAACTGATGATGGCATCTTCATCTCGGTATG 60
QY 61 GTGATGTTCTTAAGAGAGCTCTCTCTCAAGTGCATCACTTTGGAGGGCATATGCA 120
DB 61 GTGATGTTCTTAAGAGAGCTCTCTCTCAAGTGCATCACTTTGGAGGGCATATGCA 120
QY 121 TGCCAAATTTGAAGCTCTCTATCACTACCTGCTGCTGTAGCATATCCGGAGGCGA 180
DB 121 TGCCAAATTTGAAGCTCTCTATCACTACCTGCTGCTGTAGCATATCCGGAGGCGA 180
QY 121 TGCCAAATTTGAAGCTCTCTATCACTACCTGCTGCTGTAGCATATCCGGAGGCGA 180
DB 121 TGCCAAATTTGAAGCTCTCTATCACTACCTGCTGCTGTAGCATATCCGGAGGCGA 180
QY 181 TATTGAATCTTTAAGCTCTGTTAGTATGATTTCAACCTCTTATGACATTTATTTGACAG 240
DB 181 TATTGAATCTTTAAGCTCTGTTAGTATGATTTCAACCTCTTATGACATTTATTTGACAG 240
QY 241 ATACAAACGATGCTCTATGCAATCAACCTAGCAAGAGAGAGGAGGCTGATTTGGGGTGA 300
DB 241 ATACAAACGATGCTCTATGCAATCAACCTAGCAAGAGAGAGGAGGCTGATTTGGGGTGA 300
QY 241 ATACAAACGATGCTCTATGCAATCAACCTAGCAAGAGAGAGGAGGCTGATTTGGGGTGA 300
DB 241 ATACAAACGATGCTCTATGCAATCAACCTAGCAAGAGAGAGGAGGCTGATTTGGGGTGA 300
QY 301 AGTTGATCCCTGCTGAGGCGACGACAAATTTCTCAAGATGAGTACATCTTTTCGGGA 360
DB 301 AGTTGATCCCTGCTGAGGCGACGACAAATTTCTCAAGATGAGTACATCTTTTCGGGA 360
QY 361 AGTTGATCCCTGCTGAGGCGACGACAAATTTCTCAAGATGAGTACATCTTTTCGGGA 420
DB 361 AGTTGATCCCTGCTGAGGCGACGACAAATTTCTCAAGATGAGTACATCTTTTCGGGA 420
QY 421 AGTTGATCCCTGCTGAGGCGACGACAAATTTCTCAAGATGAGTACATCTTTTCGGGA 480
DB 421 AGTTGATCCCTGCTGAGGCGACGACAAATTTCTCAAGATGAGTACATCTTTTCGGGA 480
QY 481 TAAAGCGTGGTGGCAAAATTTCCAAATTTGAAGAGTGGTATGAGGAGCATGTTACACCA 540
DB 481 TAAAGCGTGGTGGCAAAATTTCCAAATTTGAAGAGTGGTATGAGGAGCATGTTACACCA 540
QY 541 TTGATGCTGTTAAGTTTGTGATCTTGACATGAAAGATTGTTGACCAACTGTACACC 600
DB 541 TTGATGCTGTTAAGTTTGTGATCTTGACATGAAAGATTGTTGACCAACTGTACACC 600
QY 601 CACAAATCTTCTTTTGAACAGGAATCTCTCTTCAAGGGGGCTTACACCGCATATTT 660
DB 601 CACAAATCTTCTTTTGAACAGGAATCTCTCTTCAAGGGGGCTTACACCGCATATTT 660

DB 601 CACAAATCTTCTTTTGAACAGGAATCTCTCTTCCAMAGGGGGCTTACACCGCATATTT 660
QY 661 ACTGCTCTTCACTCTCCCAAAAGAGATCCACAGGAGGCACTGTGTGAGCTGTACAA 720
DB 661 ACTGCTCTTCACTCTCCCAAAAGAGATCCACAGGAGGCACTGTGTGAGCTGTACAA 720
QY 721 GTGGAAGTAAAGATTTTCTTGTGAGCTGATAGTGTCTCTCATATGACGAAAGAA 780
DB 721 GTGGAAGTAAAGATTTTCTTGTGAGCTGATAGTGTCTCTCATATGACGAAAGAA 780
QY 781 AGTTGCTTGTGAGTGTCTGATTTTACATGACGACCTGTGACCTGTGATATGCA 840
DB 781 AGTTGCTTGTGAGTGTCTGATTTTACATGACGACCTGTGACCTGTGATATGCA 840
QY 841 AGTTGCTTGTGAGTGTCTGATTTTACATGACGACCTGTGACCTGTGATATGCA 900
DB 841 AGTTGCTTGTGAGTGTCTGATTTTACATGACGACCTGTGACCTGTGATATGCA 900
QY 901 CAGATTTTATGAGGCTCTCTAGGACAACTCAAGATTAAGTTAGTAAAGCCCATGGA 960
DB 901 CAGATTTTATGAGGCTCTCTCTAGGACAACTCAAGATTAAGTTAGTAAAGCCCATGGA 960
QY 961 AGGTACCCGAAATCTTTTCTTATGATGATGACGAGATATTTCCATGTTGCTGTGAA 1020
DB 961 AGGTACCCGAAATCTTTTCTTATGATGATGACGAGATATTTCCATGTTGCTGTGAA 1020
QY 1021 TGCTGACATGTTGCTGCTCTCTCTGAGAAATCATTTGATCTTGTATATAT 1080
DB 1021 TGCTGACATGTTGCTGCTCTCTCTGAGAAATCATTTGATCTTGTATATAT 1080
QY 1081 GTGATTCACCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
DB 1081 GTGATTCACCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 ATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1141 ATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 CTGACCAAGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 CTGACCAAGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 ACTGATTAATA 1271
DB 1261 ACTGATTAATA 1271

RESULT 3
AAL44159
ID AAL44159 standard; DNA, 1743 BP.
XX
XX AAL44159;
AC
XX
XX 03-OCT-2002 (first entry)
DT
XX
DE Beta vulgaris dihydroorotase gene sequence.
XX
XX Sugar beet; gene; ds; transgenic plant; osmotic stress tolerance;
KW oxidative stress tolerance; casein kinase alpha catalytic subunit;
KW dihydroorotase; translation initiation factor 1A; plant flowering;
KW plant growth stimulation.
XX
XX Beta vulgaris.
OS
XX
XX Key Location/Qualifiers
FH 199. 1236
FT CDS /*tag= a
ET /product= "Beta vulgaris dihydroorotase"
XX
XX W0200252012-A2.
XX
XX PD 04-JUL-2002.

Query Match	50.8%	Score 646;	DB 6;	Length 1743;
Best Local Similarity	76.4%	Pred. No. 1e-176;		
Matches 793; Conservative	0;	Mismatches 245;		1-2-3

QY	3	GGAAAAATGAGACTCTCAATCAACACACCTGATGATTGGACTCTTCACTCTCCGTATGG	62
Db	193	GTPAAAGATGGAACCTACCTCTTACACGCCCTGATGACTGGATCTCAATCTCCCGCATGGA	252
QY	63	GATGTTCTTAAAGGCAAGTGTGTCTTCAACAGGCAACATACCTTTGGAGGGCAATATGCTATG	122
Db	255	GATCTTCTTGCTGCTGTTTGTCTCTTCAACAGTGCAGAAACATTTTGGAAAGGGCAATCGTTATG	312
QY	123	CCAAATTTGAAAGCCTCCTATCACTACACACTGCTGCTGCTGTAGACATACCGGGAGGGAGTA	182
Db	313	CCGATCTTAAGGCCCTCGTGTACTACTACAGAGTGCTGCTATTTGCTTACCGAAAGTCTAT	372
QY	183	TTGAAATCTTAACCTGTTGATAGTGAATTTCAACCCCTTTATGACACTTTATTTGACAGAT	242
Db	373	ATGGAAGTATTTGCTGATAGTAGAAGGAACTTCAATCCCTCATGACACTTATTTGACTAT	432
QY	243	ACAAACAAGTCTATGAAATCAAACTAGCAAGAGAGAGCAGTGCATTTGGGGTGAAG	302
Db	433	ACGACCAAGCCCTAATGAGATCAAGCTTCCAGAAAAAGAGAGTGGTATATGCTCTCAAA	492
QY	303	TTTGAACTCTGCTGTGTCACGACAAATTTCTCAAGATGAGTACTGATCTTTTGGGGAAG	362
Db	493	TTTAACCTCTGCGCGGACAACTAATTTCTCAGAGACGGTGCATGATCTTTTAGAAGAG	552
QY	363	TGTTTACAGATTTCTACAGAAATGAGTTGAGCAATATATGCTCTGCTGTTCATGGAAG	422
Db	553	TGCTGCTGTGCTTTGAAGAGATGGCTGAGCAAGATATGCTCTTCTGTGCATGAGAA	612
QY	423	GTTTACTAATCTGAGGTTGACATGTTTGAATAGAAAAAGTATTCATTGAAACGGTCTTA	482
Db	613	GTTACAGATCTCTGATGATATATTTGATTCGTGAAGAGTTTTTATTAGATCAAGTTTAA	672
QY	483	AGACCGTGGTAGAAATTTTCCACAATTTGAAAGTCTGATGAGAGCATGTTACCAACAT	542
Db	673	AGACCTTTAATTCAGAAATTTACACAGCTTAAAGGTTGTGATGAGAAACATCATCACTGCT	732
QY	543	GATGCTTTAAGTTTGTGTAATCTTGCACTGAAGATTTGTTGAGACACTGTCAACCCA	602

Db 733 GATGCTGCMAAGTTTATTGAGTCTCTGTAATGGAAGAAATGATGACGCCATGTCACCCG 792

Qy 603 CAACATCTGTGTTTGAACAGAAATTCCTCTTTCGAAGGGGCTTACAACCGCATAAATTAC 662

Db 793 CAGCACCTGTGTTGAATAGAACTCTCTTCCAAAGAGGGTTGCAACCGCATTAAT 852

Qy 663 TGCCTTCAGTCTCAAAAAGAGATCCACAGAGGGACCTGTGTGATGCTTAACAAT 722

Db 853 TGTCTTCAGTGTCAAAAAGAAATCCATAGACAGCATTTGTTTACGGGTACCAGC 912

Qy 723 GGAATGAAGAATTTTTTCTGGACTGATATGTCTCTCTCATGATAGCAGAAAAAGAG 782

Db 913 GGGAGCAGCAATATTTTCTTGAGACTATATGCTCTCATGAAAAGCGAGAAAGAA 972

Qy 783 TGTTCCTTGAGATGTGCTGTAATTTACATATCACCCTGAGCCTGTGTAGATATGGGAG 842

Db 973 TGTTCGTGTGATGTCTGGAATCTATATTTCCCTGTGTCTATCATATATGCGAAA 1033

Qy 843 GTGTGTAAGAAAGSAAATGCACTGACATGCTTGAACATTCATGACTTCAATGACCA 902

Db 1033 GTATTTGAAAGAGCTGTGCTGCTTCACAAAGTTAGACGATTTCAAGCTTTATGACCC 1092

Qy 903 GATTTTATGAGCTTCTAGGAACAACATAAGATTAAGTTGATGAAGAGCCATGGAG 962

Db 1093 GATTTCTATAGCTTCTTCCAAAGAAATGCGTCAAGATCAAGTTGAAAAAAGAACCTAGGAAA 1155

Qy 963 GTACCCGAATCCTTTTCTATGCAATCAGAGATATTTATTTCCATGTTTCTGTGAAATG 1022

Db 1155 GTCTTAGAGCGATATCTTTTCCCATTCGGAGAAATATATCCTATGTTTGTCTGACAAATG 1212

Qy 1023 CTCGACTGTGTCGGGCT 1040

Db 1213 CTTGACTGGAAGCCATCT 1230

XX	ACAA53620
XX	ACAA53620 standard; DNA, 1053 BP.
XX	ACA53620;
XX	19-JUN-2003 (first entry)
XX	Prokaryotic essential gene #35277.
XX	Antisense; ds: prokaryotic essential gene; cell proliferation;
XX	drug design; gene.
XX	Vibrio cholerae.
XX	WO200277183-A2.
XX	03-OCT-2002.
XX	21-MAR-2002; 2002MO-US009107.
XX	21-MAR-2001; 2001US-00815242.
XX	06-SEP-2001; 2001US-00948993.
XX	25-OCT-2001; 2001US-0342923P.
XX	08-FEB-2002; 2002US-00072851.
XX	06-MAR-2002; 2002US-0362699P.
XX	(ELIT-) ELITRA PHARM INC.
XX	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
XX	Mall D, Trewick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
XX	WPJ; 2003-029926/02.
XX	P-PSDB; ABU49750.
XX	New antisense nucleic acids, useful for identifying proteins or screening
XX	for homologous nucleic acids required for cellular proliferation to

PT Isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 41490; 1766bp; English.

XX
PS
XX The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1053 BP; 267 A; 299 C; 257 G; 230 T; 0 U; 0 Other;

Query Match 22.1%; Score 281; DB 8; Length 1053;

Best Local Similarity 56.8%; Pred. No. 9.8e-71;

Matches 578; Conservative 0; Mismatches 430; Indels 9; Gaps 3;

QY 15 CTCTCAATCACAACCTGATGATGGCATCTTCATCTCCGTATGCTGATGTTCTTAAG 74
DB 34 CTGACGATTACCGCCGACGACGACGACGACGACGACGACGACGACGACGACGACG 93
QY 75 GCACTGTCTCTCAAGTCACATCACTTTGGAGGGCAATAGCATGCGCAATTTGAA 134
DB 94 GATACGTAGCGCATCAGCCGTTATACGCGCGCGCTGATCATGCTTAACACGGTT 153
QY 135 CCTCCATCATCAACCTGCTGCTGTAGCATATCCGAGGCGATTTGAAATCTTTA 194
DB 154 CCCCTGTAAACACACGATGATGCGCTTACGCTATCGTGACGATTAATGCGCGTAG 213
QY 195 CCTGTGTAGTATGATTCAACCTCTTATGACATTTATTTGACAGATCAACAGTCTT 254
DB 214 CC--ACAAAGCGCACTTTGAACCTTTATGCGGCTTTTACCTACCGATTAACATCACT 270
QY 255 ATGGAATCAACATTAAGCAAGAGAGCGAGTCGATTTGGGGTGAATTTGACCCGCT 314
DB 271 GAGAAATTCGCAAAAGCCAAAGCGTCAGGCAAAAGCTTGAGCTAACTTTAACCCACC 330
QY 315 GGTGCGACGACAAATTTCTCAAGATGAGTGACTGATCTTTTCGAGAAAGTGTTCACAGTT 374
DB 331 GGTGCGACACCAACTCGGATTCAGGCGTCACTTCAAGCAAAAACATTTACCACTT 387
QY 375 CTACAAAGATGTTGAGCATATATGCTCTGCTGCTTCAATGAGAGGTTATCTAATCTT 434
DB 388 TTGCAAGGATGACAGAAAGTCGATGTTGCTGTGGTGAACGCGGAAGTCAACCGAC 447
QY 435 GAGGTTGACATGTTGATGAGAAAGATTCATTGAAGGATCTTAAGACCGTGGTG 494
DB 448 GAAGTCGACATTTTGAACCGTGAAGAACCTTCTAGACACCGTCTTGCGCGCATGCTG 507

QY 495 CAGAAATTTCCAAATTTGAAGTCGATGAGACATGTTACCAACATGATGCTGTTAAG 554
DB 508 AATGATTTCCCGACACTGAAAATTTGTCTTGAACATCAACACCCCGATGAGTGAAC 567
QY 555 TTTGTTGAATCTTGCACATGAGATTTGTTGACGACATGTCACCCCAACATCTTGT 614
DB 568 TTTGTGCA--CAACAGCGGATTAAGTTGGCGCAACATTAACCGCGACCATTTGCTG 624
QY 615 TTGAACGAGATTTCTCTTCCAAAGGGGCTTCAACCCGATTAATTAATGCTTCCAGTC 674
DB 625 TTAAACCGTAAACCAATGCTGCTGGTGGCGGATTTGCCCATTTCTAATGTTGCCAATC 684
QY 675 CTGAAAGAGATTCACAGGAGGAGCACTTGTGCTGACGTGTAACAAGTGAAGTAAAGA 724
DB 685 CTAAACGTCGACCTACCAACAGCGGTTATGCGAGCGGCACTTGGACAGAGAAA 744
QY 735 TTTTCTTGGGACATAGTGTCTCTCATGATGACGAAAGAGTGTCTTGTGA 794
DB 745 TTCTTCTTAGGTACGACACTGCTCCCGACGCAAAAGCCGCAAGAGCGCTGTGTGC 804
QY 795 TGTGCTGTATTTTACATGACCTGTAGCCTTGTCAATATGCAAGGTTTGAAG 854
DB 805 TGGCAGGTTCTTACACAGCCCATGACGCTTGAATGTATCCGAAGTGTGAAAA 864
QY 855 GAAATGCACTGACAAAGCTTGAAGCATTCATAGCTTCAATGAGCAACGATTTTATGG 914
DB 865 GAAAGTAAAGCTAAGAAATCTGAAGGCTTGTGCTTGAATGAGCCCTGATTTCTACGGC 924
QY 915 CTTCCTAGAAACAATCAAGATTTAAGTGAATGAGACGACATGAGAGGTAACCGAATCC 974
DB 925 CTACCTGCAATGAAGACCGTCAACGCTCAACCAAGAACCTGGCCAGTGGCAAGAAC 984
QY 975 TTTTCTTATGATCAAGAGATTTATTTCCATGTTTCTGCTGAATGCTGACTGG 1031
DB 985 ATGCCGTTTGGCAGCATATCGTGTGTCATTCGCGCCGCGTGAATAATGAGTGG 1041

RESULT 5
ADEB9827
ID ADEB9827 standard; DNA; 1071 BP.
XX
AC ADEB9827;
XX
DT 12-FEB-2004 (first entry)
XX
DE Bacterial polynucleotide #113.
XX
KM Proteus mirabilis infection; bacterial infection; antibacterial;
KM Immunostimulant; gene; ds.
XX
OS Proteus mirabilis.
PN US6605709-B1.
XX
XX 12-AUG-2003.
PD
XX 05-APR-2000; 2000US-00543681.
PF
XX 09-APR-1999; 99US-0128706P.
PR
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
PI Breton GL;
XX
XX WPI; 2003-895291/82.
DR P-PSDB; ADF04000.
XX
PT New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.
XX

or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPD at ftp.wipo.int/pub/published_pct_sequences

Sequence 1047 BP; 249 A; 282 C; 237 G; 279 T; 0 U; 0 Other;

Query Match 21.8%; Score 277.2; DB 8; Length 1047;
Best Local Similarity 55.9%; Pred. No. 1.2e-69;
Matches 569; Conservative 0; Mismatches 443; Indels 6; Gaps 2;

```
QY 15 CTCTCAATCAGACAACTGATGATGGCATCTTCATCTCCGATGATGATGTTCTTAAG 74
DB 22 CTAAATTCGCGCGCCGAGTACTGGACATTCATCTAGTGATGTAAGTAATGCTCAGT 81
QY 75 GCAGTTGCTCTCAGTGCAGTGCATCATCTTGGAGGAGCAATAGCATGCCAAATTTGAG 134
DB 82 ACCGTGTTGCCCTTACCTCGAGATATTCGCCGCTTATGTTATGCCAAATCTAGCC 141
QY 135 CCTTCATCACTACACTGCTGCTGTAGCATATACGGAGGAGCATATTGAATCTTTA 194
DB 142 CAGCATTTAAAGGTTGCGAGTACTATTGCTATCGGAGCGTATTTTAAAGCAGCGTT 201
QY 195 CTTGTTATGATGATTTTCAACCTCTTATGACACTTTATTTAGACAGTACACAGTCT 254
DB 202 CTTGCGGCGCATTAATTCACCCCGTTGATGACGTTTACGACTTAATAGCCTTGAATGCT 261
QY 255 ATGGAATTCAACTAGCAAGAGAGAGGAGGAGGAGTCTATTTGGGAGGAGTGTACCTGCT 314
DB 262 AAAGAGTTGACCAAGGCTTTTGAAGCAAGGCTTTTACCGGAGCAACGTATACCGGCC 321
QY 315 GGTGCAAGCAAAATTCCAAGATGATGATGATCTATTTTCCGGAAGTGTTCACAGTT 374
DB 322 AATGCAACCAACCACTCACTCAGCGTGTATCTGACAT---CCGGCAATTTACCGCGT 378
QY 375 CTACAGAAATGTTGAGCATATATGCTCTGCTGTTCTATGAGAGGTTACTAATCCT 434
DB 379 TTGAAACAAATGCAAAAGATAGGATGCGCTGCTTATTTCAAGGTGAGTAAAGATGCG 438
QY 435 GAGGTGACATGTTTGTATGAGAAAGATTTTCAATTGAAAGGTTCTTAAAGCCGTTGGTG 494
DB 439 GCCGTGACATCTTGTATGCGAAGCCGTTTATTTAGCAAAATTTTGAAGCCATTCGCG 498
QY 495 CAGAAATTTCCAAATTTGAGAGTGTGATGAGAGATGTTACACCATTTGATGCTGTTAAG 554
DB 499 CAAAGTTTCCGAACTAAATTTGCTTTGAGCATATACGACCAAAAGATGCGGAGAT 558
QY 555 TTTGTTGATCTTGTGACATGAAAGATTTTGTGAGCAATGTACACCCCAACATCTTGT 614
DB 559 TATGTGTCGAGGGA--ATCGTTTCTTGGGGGAAACGTCACGCCCAACACTTGATG 615
QY 615 TTGAACAGGAATTTCTCTTCCAAAGGGGCTTAAACAGCAATTAATTAATGCTTCAAGTC 674
DB 616 TTTAACCGCAATTCACATCTGTGAGCGGTATTCGCCCCCACTTGTTCCTGCGAATA 675
QY 675 CTCAAAAGAGATTCACAGGAGGAGCACTTGTGTCAGTGTGTAACAGTGAAGTAAGAA 734
DB 676 TTGAAGCGAGCAGCATCAGCAAGCATTTGCGCGAGCGGTGCGCAGGTGTTGATGCG 735
QY 735 TTTTTCCTTGGAGATGATGCTCTCTATGATTAACGAAGAAAGATGTTCTTGGAGA 794
```

```
DB 736 TTCTTCCTTGGAGCAGGATTCAGCTCCCATGCCAAACATGTAAGATCATCTTGGCGC 795
QY 795 TGTGCTGATTTTAAATCATTAGCTTGTCACTATATGCGAAGGTGTTGAAAG 854
DB 796 TGTGCGGGGTATTTCAACCCCGAGGAGATTCGCTTAAATGCTTCCCTGTTGGAA 855
QY 855 GAAATGCACTGCAAGAGTTGAGCATTCAGTACCTCAATGACACAGATTTTATGGG 914
DB 856 CTGAATGCAATTCACATCTGGAAGCGTTTGCGCTTAAATGCGCCAGATTTATGCG 915
QY 915 CTTCCTAGGACAACTCAAGATTAAGTTAGTAAAGCCCATGAGGATCCCAATCC 974
DB 916 TTGCGTGAATGATGACGTTGTAATTTGTCGACATCCATTCCTGACAGCCAGAAAG 975
QY 975 TTTTCTTATGATCAGAGATATATTTCCCATGTTGCTGTGTAATGCTCGACTG 1032
DB 976 ATCCCATTAAGCAATGAATCGGTATTCCTTTCCTTGGCGGTCAACGCTTAATG 1033
```

RESULT 7

ACF68579
ID ACF68579 standard; DNA; 1053 BP.

XX ACF68579;

DT 20-NOV-2003 (first entry)

DE Photobadus luminescens nucleotide sequence #7046.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;

KW detection; food; gene expression; plant; animal; microorganism; toxin;

KW antibiotic; biopesticide; virulence factor; disease model; plague;

KW whooping cough; gene; ds.

OS Photobadus luminescens.

PN NO200294867-A2.

PD 28-NOV-2002.

PF 07-FEB-2002; 2002MO-IB003040.

PR 07-FEB-2001; 2001FR-00001659.

PA (INSP) INST PASTEUR.

PA (CNRS) CNRS CENT NAT RECH SCI.

PI Duchaud B, Taourit S, Glaeser P, Frangeul L, Kunet F, Danchin A;

PI Buchrieser C;

DR WPI; 2003-148459/14.

XX Genomic sequence of Photobadus luminescens and encoded polypeptides,

PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

PS Claim 2; SEQ ID NO 7046; 1205BP; French.

XX The invention relates to the isolation of genes and their encoded

CC proteins from Photobadus luminescens. The isolated sequences are

CC sources of probes and primers for detecting the genome of *P. luminescens*

CC and related species; to study polymorphisms; for gene analysis and for

CC detection/amplification of the genes. Antibodies (Ab) raised against the

CC polypeptides encoded by the genes are used for detection/identification

CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that

CC carry a gene-containing vector are used to select compounds that

CC modulate, regulate, induce or inhibit expression of the genes in plants,

CC animals or microorganisms other than *P. luminescens* and are able to alter

CC response or sensitivity to toxins and antibiotics produced by *P.*

CC *luminescens*. Cells transformed to express the genes are useful for

CC recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The

CC genes, proteins, vectors containing the genes and Ab are also useful

CC therapeutically (to treat microbial infection by bacteria or fungi that

20 AATCAACAACCTGATGATGGCAATCTTCATCTCCGTGATGGATGATGTTCTTAAGCAGT 79
27 AATCGCCCCCTGATGACTGGCAATCAATCTTTCGATGATGATGAATGTAAGAACTGT 86
80 TGTCTCTCAAGTGCATCACTTGGGAGGGCAATGTGATGCGAAATTTGAAGCTCC 139
87 CGTTCCTTACACGATCATTAATTTGGCCGAGCTATGTCATGCTTAATCTGTTCAGCC 146
140 TATCACTACCACTGTGCTGCTGTATGCAATACCGGAGGCGAATTTGAAATCTTACTGT 199
147 GATTCAGAAAGTGTCACTGCGCAAGGCTTTAAGAGGCAAGAACTGGCAGCTAATCCGA 206
200 TGAATGTGATTTCAACCTCTTATGACACTTTATTTGACAGATATACCAACGCTTAATGA 259
207 AGGAGATATTTTCCAGCTCTTATGACTTGTATCTTATCTGACACGACTGAAAGCTCTCA 266
260 AATCAAACTACAGAGAGAGCCAGCTGTAATTTGGGGTGAAGTTGTAACCTCGTGTGC 319
267 GATGAAATTTGTTATTAAGAGGCAATTTTACCGTTTGCAAGCTCTACCTCGCAATGC 326
320 CACGACAAATTTCTCAAGATGGAATGATCTGTTTCCGGAAGTGTATACAGTTCTACA 379
327 CACAACCAATTCAGTCAAGCTCAGCGCTCTCTGATATAA--GAAATTTATCCAATATAGC 383
380 AGAATGTTGAGCATATATATGCTCTGCTGCTGTTTACGAGAGTTACTTAATCTTGAGT 439
384 CGTATGGAATAACTGGGCAATGCTCTTGTCTGTGTCACGATAGGTTATGCTTCTCATAT 443
440 TGAATGTTGATGAGCAAAAGGATTCATTGTAACCGGTTCAAGCCGTTGGTGCGAAT 499
444 TGAATTTTGATGAGGAAGCCCGTTTATTTAACAATCATGGAACACTAGTAATCA 503
500 ATTTCACAAATTGATGAGTGTGATGAGAGATTTACCAACAATTGATCTGTAAATTTGT 559
504 ATTCACGAGCACTGAATGTTGTGTTTGAACACATTACTACCAAAAGCCGACACAATATGT 563
560 TGAATCTGACATGAAGATTTGTTGCAAGCACTGTCAACCCACAAATCTGTTTGA 619
564 GCTGGAAGGCAATA--CAATCTGGCAGCAAACTTAATCTCACAGACTCATATGTTCA 620
620 CAGGAATCTCTTTCGAAGGGGCTTTCACAACCGCATTAATTAATGCTTCCAGTCTCA 679
621 TCGCAACCAATGATGTAATGTGGGGAATTCGGCCACAATTAATCTGTCTGTGTGTAA 680
680 AAGAGAGATCCACAGGGAGGCACTTGTGTCACTGTATAACAGTGAAGTAATAAGATTTT 739
681 AGCCAATGTGATCAAGAGGCTTTTACGGGACGCGTGCAGAGTGCGTGTGATCGCTT 740
740 TCTTGGGACTGATAGTGTCTCTCATGATGATGACGAAGAAAGATGTTCTTGGGATGTGC 799
741 TCTTAGTACAGACTCAGCACCAATGCAACAAGAAAGAAAGATCCTCTTGTGTGTGTGC 800
800 TGGTATTTACATGCACTGTAAGCTTGTCAAGTATATGGAAGGTTTGAAGAAGAAA 859
801 CGGCGCTCTTAAACGACATCAGGCGCTAATCTGTTATGCAACAGGTTTGAAGAGATGA 860
860 TGCATCGACAAGCTTGAAGCATTTCACTAGCTTCATGACCAAGATTTTAAATGGCTTCC 919
861 TGCAATTACCGCATTTTGAAGCTTTCTGTTCATTTAATGACCTTAATTTTTCGGGCTACC 920
920 TAGGAACAACATGAAGTTAAGTTGAGTAAAGCCAGTGAAGGTAACCCGATCTTTTC 979

WP	ACF67367_00	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5648894

Query Match

Best Local Similarity

21.5%;

55.7%;

Score 272.8; DB 10;

Pred. No. 2.1e-67;

Length 110000;

Matches	564;	Conservative	0;	Mismatches	442;	Indels	6;	Gaps	2;
Qy	20	AATCACAACCTGATGATGGCATCTTCATCTCGGTGATGATGTTCTTAAGGCAGT	79						
Db	43973	AATCCGCCGCCGATGATGATGGCATCTTCATCTCGGTGATGATGTTCTTAAGGCAGT	44032						
Qy	80	TGTCTCTCACAGTGCATCATCTTTGGGAGGCGCAATGTCACCAATTTGAAGCTCC	139						
Db	44033	CGTTCCTTACACAGTGCATCATCTTTGGGAGGCGCAATGTCACCAATTTGAAGCTCC	44092						
Qy	140	TATCATAACACCTGCTGCTGTAGCATCCGAGGAGGCAATTTGAATCTTACCTGT	199						
Db	44093	GATTACAGAAAGTGTAGTGCAGAGCTTTAAGGCGAAATCTGCGACTTTTCCCA	44152						
Qy	200	TGATAGTATTCACCTCTTATGACACTTTATTTGACAGATACACAGCTCTATGA	259						
Db	44153	AGGAGATATTTCCAGCTCTTATGATGTTATCTTACTGACAGCTGAAGCTCTCA	44212						
Qy	260	AATCAAACTAGCAAGAGAGACCGAGTCTGATTTGGGAGTAAATTTGACCTCTGTC	319						
Db	44213	GATAGAAATGGTTATTAAGAGGCAATTTTACCGCTTGCAAGCTTACCTGCAATGC	44272						
Qy	320	CACGACAAATTTCAAGATGAGATGATGATCTTTTGGGAGTGTACCAATCTTACA	379						
Db	44273	CACACCAATTTCAAGTACGCGCTTGTATATAA---GATATTTATCTCACTATAGC	44329						
Qy	380	AGAAATGTTGAGCATTAATGCTCTGCTGTTTCAATGAGAGGTTACTAATCTGAGT	439						
Db	44330	CGTATGAGAAATTTGGGAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	44389						
Qy	440	TGACATGTTGATGAGAAAGATTAATCAATGAAACGTTCTTAAGACCTTGGTGCAG	499						
Db	44390	TGATATTTGATGAGAAAGATTAATCAATGAAACGTTCTTAAGACCTTGGTGCAG	44449						
Qy	500	ATTTCCACATTTGAGTGGTGGATGAGCATGTTACACCATTTGATGATGTTAGTTGT	559						
Db	44450	ATTTCCACATTTGAGTGGTGGATGAGCATGTTACACCATTTGATGATGTTAGTTGT	44509						
Qy	560	TGAATCTTGCATGAGATTTGTTGAGCACTGTGACCCACACACATCTTTGTTGAA	619						
Db	44510	GCTGGAAGGCAATA---CAATGCGAGCAACATTAATCTCAAGATCTCATGTTCAA	44566						
Qy	620	CAGAAATCTCTCTCCAAAGGCGCTTAAACCGCATTAATTAATGCTCTTCAAGTCTCA	679						
Db	44567	TCCGACCAATGCTAGTGGTGGATTCGCGCACACCTTAATGCTCTGCTGCTTAA	44626						
Qy	680	AAGAGATTCACAGGAGGACCTTGTCTCAGCTGTAAACAGTGAAGTAAAGATTTT	739						
Db	44627	ACGCAATGTGATCAAGAGGCTTTACGCGCGCGCTGCGAGTGTGATGCTCTTT	44686						
Qy	740	TCTTGGGACTGATAGTGTCTCTCATGATAGACGAAGAAAGATTTCTTGTGATGTC	799						
Db	44687	TCTTGGGACTGATAGTGTCTCTCATGATAGACGAAGAAAGATTTCTTGTGATGTC	44746						
Qy	800	TGATATTTACATGACACCTGTAGCTTGTGATAGTGAAGGTTGTTGAAAGAA	859						
Db	44747	CGGCGTCTTTAAGCGACCATGAGGCTTACTGTTATCAACAGTGTGTTGAGGAGTAA	44806						
Qy	860	TGCACTGACAAAGCTTGAAGATTCATAGCTTCAATGAGACCAATTTTATGAGCTCC	919						
Db	44807	TGCACTGACAAAGCTTGAAGATTCATAGCTTCAATGAGACCAATTTTATGAGCTCC	44866						
Qy	920	TAGAACAACTCAAGATTTAAGTGAAGACCGCATGAGAGTACCGGATCTTTTC	979						
Db	44867	TGTCAATTAAGGCTTTATTTGAACATACCTGCAAACTTCTACCGCATTTGAACATTTGA	44926						
Qy	980	TTATGATCAGAGATTAATTTATTTCCATGTTGCTGTGAAATGCTGCACTGG	1031						
Db	44927	TTGCAATTAATGAAATTAATTTATTTGCTGTGCGAAGATGCGCGCTGG	44978						

RESULT 9
ACF65379/c

ID	ACF65379 standard; DNA; 182624 BP
XX	XX
AC	ACF65379;
XX	XX
DT	20-NOV-2003 (first entry)
XX	XX
DE	Photobhabdus luminescens nucleotide sequence #32.
XX	XX
KW	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW	detection; food; gene expression; plant; animal; microorganism; toxin;
KW	antibiotic; biopesticide; virulence factor; disease model; plague;
KW	whooping cough; gene; ds.
OS	Photobhabdus luminescens.
XX	XX
PN	W0200294867-A2.
XX	XX
PD	28-NOV-2002.
XX	XX
PF	07-FEB-2002; 2002W0-1B00340.
XX	XX
PR	07-FEB-2001; 2001FR-00001659.
XX	XX
PA	(INSP) INST PASTEUR.
XX	XX
PI	(CNRS) CNRS CENT NAT RECH SCI.
XX	XX
P1	Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst F, Danchin A;
PI	Buchrieser C;
XX	XX
DR	WPI; 2003-148459/14.
PT	Genomic sequence of Photobhabdus luminescens and encoded polypeptides,
XX	XX
PT	useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX	XX
PS	Claim 1; SEQ ID NO 32; 1205bp; French.
XX	XX
CC	The invention relates to the isolation of genes and their encoded
CC	proteins from Photobhabdus luminescens. The isolated sequences are
CC	sources of probes and primers for detecting the genome of P. luminescens
CC	and related species; to study polymorphisms; for gene analysis and for
CC	detection/amplification of the genes. Antibodies (Ab) raised against the
CC	polypeptides encoded by the genes are used for detection/identification
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC	carry a gene-containing vector are used to select compounds that
CC	modulate, regulate, induce or inhibit expression of the genes in plants,
CC	animals or microorganisms other than P. luminescens and are able to alter
CC	response or sensitivity to toxins and antibiotics produced by P.
CC	luminescens. Cells transformed to express the genes are useful for
CC	recombinant production of the proteins, particularly toxins and
CC	antibacterials useful as insecticides, bactericides and fungicides. The
CC	genes, proteins, vectors containing the genes and Ab are also useful
CC	therapeutically (to treat microbial infection by bacteria or fungi that
CC	are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC	biopesticides. Other uses of the genes and the proteins are as virulence
CC	factors and for identifying targets of human diseases for which P.
CC	luminescens is a model (particularly plague and whooping cough). This
CC	sequence represents one of the isolated P. luminescens genes
XX	XX
SO	Sequence 182624 BP; 52367 A; 35105 C; 41128 G; 54021 T; 0 U; 3 Other;
Query Match	21.5%; Score 272.8; DB 10; Length 182624;
Best Local Similarity	55.7%; Pred. No. 2.7e-67;
Matches	564; Conservative 0; Mismatches 442; Indels 6; Gaps 2;
Qy	20 AATCACAACCTGATGATGGCATCTTCATCTCGGTGATGATGTTCTTAAGGCAGT 79
Db	77563 AATCCGCCGCCGATGATGATGGCATCTTCATCTCGGTGATGATGTTCTTAAGGCAGT 77504
Qy	80 TGTCTCTCACAGTGCATCATCTTTGGGAGGCGCAATGTCACCAATTTGAAGCTCC 139
Db	77503 CGTTCCTTACACAGTGCATCATCTTTGGGAGGCGCAATGTCACCAATTTGAAGCTCC 77444
Qy	140 TATCATAACACCTGCTGCTGTAGCATCCGAGGAGGCGATATTTGAATCTTAACTGT 199

OS Actinobacter baumannii.
XX
PN WO200271783-A2.
XX
XX 03-OCT-2002.
PD
PF 21-MAR-2002; 2002WO-US009107.
PR
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00946992.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
PA (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KI, Zykind JW,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-0292926/02.
DR P-Patent: ABU16776.
DR
DR New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PS isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 8516; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format.

[illegible]


```
OY 147 ACCACTGCTGCTGCTGATACCGGAGGCGATTTGAAATCTTTACTGTTGATGT 206
DB 142 ACAGTAGAAGAGCTTAGCTTATCGGAAACGATCTTGCTCATGTTCCGAGAGCAAT 201
OY 207 GATTTCAACCCCTTTTGAACCTTTATTTGACAGATCAACAGCTCCTTGAATCAAA 266
DB 202 AATTTGACCCCTGATGCTGCTTATTTATTTGACACACTTCAACAGATGAAGTTGCT 261
OY 267 CTAGCAAGAGAGAGCGAGTCTGATTTTGGGGTGAAGTTTACCCCTGCTGAGCAGACA 326
DB 262 AAAATTAAAGAACAGAACTGTAATGCGATTATAGCTTATCTGCTGGTGCGACACACA 321
OY 327 AATTTCAAGATGAGTACTGATCTTTTGGGAGTGTTCACAGTTTCAACAAGAAATG 386
DB 322 AACTCCGATTAATGCTGAGTGAATATTT---CGTAAAGTTTATGACAGTCAATTGACAAATTA 378
OY 387 GTTGAAGCATTAATGCTGCTGCTGCTGATGAGAGAGTTACTAATCCTGAGGTTGACATG 446
DB 379 GAAGAGCATCAAGTTCCGTTATGCTTCATGTAAGTAATCAACAATCATGTAGATATTT 438
OY 447 TTTGATAGAGAAAGGTAATCATTTGAAACGGTCTTAAGACCGTTGTCAGAAATTTCCA 506
DB 439 TTTGATGCTGAAAAACGCTTCTAGATGAGGTATTAACCTCTATTGAAACAGTTCCG 498
OY 507 CAATTGAAGTCTGATGAGCATGTTACCAACCATGATGCTGTTAAGTTGTTGAATCT 566
DB 499 AAATTTAAAGTTGCTGAGACATCATACAGTATGACAGACACTTTGTT---TTA 555
OY 567 TGCACTGAAGATTTGTTGACAACTGTCAACCCCAACAACATCTGTTTGAACAGGAT 626
DB 556 GAACAAAGACCGTAATGAGGCGCAACAATCATCAACAACATTAATTAATTAACCGTAAT 615
OY 627 TCTCTCTTCAAGGGGCTTACAAACCCGATTAATGCTTCCAGTCTCAAAAGAGAG 686
DB 616 GATATGTTGTTGTTGATTAATTAACCGCACTTTTACTGTTTAAACAAATTAACGTCAA 675
OY 687 ATCCACAGGAGGACCTTGTGTGCTGATTAACAAGTGAAGTAAAGATTTTCTTGGG 746
DB 676 ACACATGAACGATTTGCTTGAAGTGGCGCAAGCGGTAAATCTTAATTTTCTTGGT 735
OY 747 ACTGATATGCTCTCTCATGATGACGAAGAAAGAGTGTCTTGTGATGCTGCTGATTT 806
DB 736 ACACACAGCGCTCTCTCATGCAACAAATGCAAAAGAGATGCTTGTGATGTCAGGCTGC 795
OY 807 TACATGACACTGATGAGCTTGTGCTGATTAAGCGAAGGTTTGAAGAAAGAAATGCACTC 866
DB 796 TATAGTGACCAAAATGCAATGAGCTTTATGCAACAGATTTGACAGGTGCGTAATTA 855
OY 867 GACAGAGCTTGAAGCATTCATAGCTTCAATGAGCCAGATTTTATAGGCTTCTTAGAAGC 926
DB 856 GAGGCTGTGAAGAGTTTGTGATGCAATTTTGTGTGGGACTTCTAGCGTCTACACAGTAAT 915
OY 927 AACTCAAGATTAAGTGAAGTAAGACGCGCATGAAAGGTACCCGAATCTTTTCTTAT 983
DB 916 ACTCTACCATTAATCTTGTGTTAAAGAAAGATAACCTCGTTCCAGAAATCTTTGATTAAT 972
```

RESULT 11
ADA30128 standard; DNA; 1041 BP.

ADA30128;

20-NOV-2003 (first entry)

DNA encoding *Acinetobacter baumannii* protein #1415.

de; gene; *Acinetobacter baumannii*; bacterial disease; antibacterial;
vaccine; plant biocontrol agent.

Acinetobacter baumannii.

XX

```
PN US6562958-B1.  
PD 13-MAY-2003.  
XX  
XX 04-JUN-1999; 99US-00328352.  
PF  
XX 09-JUN-1996; 98US-0088701P.  
PR  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA  
XX Breton G, Bush D;  
PI  
XX WPI: 2003-576092/54.  
DR P-PSDB; ADA34254.  
DR  
XX  
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents  
PT for diagnosing a bacterial disease, as components of antibacterial  
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for  
PT plants.  
XX  
XX Example; SEQ ID NO 1415; 328bp; English.  
PS  
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.  
XX The A. baumannii nucleic acids and polypeptides are useful as reagents  
CC for diagnosing a bacterial disease, as components of antibacterial  
CC vaccines, as targets for antibacterial drugs, to detect the presence of  
CC A. baumannii and other Acinetobacter species in a sample, in screening  
CC compounds for the ability to interfere with the A. baumannii life cycle  
CC or to inhibit A. baumannii infection, and as biocontrol agents for  
CC plants. The present sequence represents DNA encoding an A. baumannii  
CC protein.  
XX  
SQ Sequence 1041 BP; 313 A; 217 C; 204 G; 307 T; 0 U; 0 Other;  
XX  
Query Match 21.0%; Score 267.4; DB 9; Length 1041;  
Best Local Similarity 56.4%; Pred. No. 8.8e-67;  
Matches 540; Conservative 0; Mismatches 411; Indels 6; Gaps 2;  
OY 27 CAACCTGATGATTTGGCATCTTTCATCTCCGATGATGATGTTCTTAAGGAGGATGCTCT 86  
DB 28 CAGCGAATGATTTGGCATGACATTTACGATGATGATTTAGCATTAAGAAACGTAATGTTCA 87  
OY 87 CACAGTCACATCATCTTTGGGAGGCAATATGTCACCAATTTGAAGCTCTATGACT 146  
DB 88 GATTTGGCTAAACATTCGCCCGCCGATTTGATGCTTAACCTTGATACCGCTGTAAAA 147  
OY 147 ACCACTGCTGCTGCTGATACCGGAGGCGATTAATGTAATCTTAACTGTTGATAGT 206  
DB 148 ACAGTAGAAGAGCTTAGCTTATCGGAAACGATCTTGCTCATGTTCCGAGAGCAAT 207  
OY 207 GATTTCAACCCCTTTATGACATTTATTTGACAGATCAACAGCTCCTATGGAATCAAA 266  
DB 208 AATTTTGAACCTCGTATGCTGCTTATTTATTAACACACTTCAACCTGATGAAGTTGCT 267  
OY 267 CTAGCAAGAGAGAGCGAGTGTATTTGGGGTGAAGTTTGAACCTGCTGAGCAGACA 326  
DB 268 AAAATTAAAGATCAGAACATGTAATGCGATTAAGCTTTATCCCTGCTGAGCAGACA 327  
OY 327 AATTTCAAGATGAGTACTGATCTTTTCCGGAAGTGTATTAACAGTCTTCAACAAGAAATG 386  
DB 328 AACTCCGATTAATGCTGAGTGAATATTT---CGTAAAGTTTATGACAGTATGAGCAATTA 384  
OY 387 GTTGAAGCATTAATGCTGCTGCTGCTGATGAGAGAGTTACTAATCCTGAGGTTGACATG 446  
DB 385 GAAGAGCATCAAGTTCCGTTATGCTTCATGTAAGTAATCAATCAATCATGTAGATATTT 444  
OY 447 TTTGATAGAGAAAGGTAATGATTAAGAAACGGTCTTAAGACCGTGTGTCAGAAATTTTCCA 506  
DB 445 TTTGATGCTGAAAAACGCTTCTAGATGAGGTATTAATCACTCTAATTAACAGATTTCCG 504  
OY 507 CAATTGAAGTCTGATGAGCATGTTACCAACATGATGCTGTTAAGTTGTTGAATCT 566  
DB 505 AAATTTAAAGTTGCTGAGACATCATACAGTATGACAGCAACTTTGTT---TTA 561
```


DB 564 CTACGTCGCGGATCCGACCGCGGCTCAGGCGCGGATCGGCGCACCATCACCGCGCACCA 623
QY 608 TCTTGTGTTGACAGGAAATTTCTCTTCCAAAGGGGGTTTAAACACCGCATTAATTAATCTGCT 667
DB 624 TCTGCTGTACACCGCAATGCGATGTTTTCGCGGGATCGTCCACATTAATTAATCTGCT 683
QY 668 TCCAGTCTCAAAAAGAGATCCACAGGAGGACCTTGTCTCAGCTGTAACTAGTGAAG 727
DB 684 GCGGTCCTCAGCGCGGACGATCGGATCGGCTGTGAGGCGCGGACGTCGCGGCAA 743
QY 728 TAAAGATTTTCTTGGGACGATAGTGTCTCCATGATAGACGAAGAAAGAGTTC 787
DB 744 TCCGCGCTTCTCTCGGACCGACGCGCGCGACGCAAGCGCGGAAAGGAGCGCC 803
QY 788 TTGAGATGTCTGTGTTTATTAATGACCTGTAGCTTGTAGTATATGCGAAGGTGT 847
DB 804 GTGCGGCTGCGCGGCTGCTACACCGCGCTGCAACGCGCTGAGCTGTACCGGAGGCAAT 863
QY 848 TGAAGAGAAATGCACTGACAAAGCTTGAAGCTTCACTAATGATGACCAATTT 907
DB 864 CGACAGGCGGCGCGCTCGACAAAGCTTGAAGCTTCTTCTGCGCGGACCTT 923
QY 908 TTATGGCTTCTTACAGCAACCTCAAGATTAGTGAAGACGCGCATGGAAGTACC 967
DB 924 CTACGCTTTCGCGGACCGCGGACGAGTGAAGCTGCGCGGACGAGTGAAGTGC 983
QY 968 CGAATCTTCTTATGATCAGAGATATATATCCATGTTGCTGTGTAATGCTGCA 1027
DB 984 GCGGAGATGACCGCGGCGCGCGCGCGCTGCTGCGCGGCGGCGGAGGATCGG 1043
QY 1028 CTGCTTGC 1035
DB 1044 CTGCGCGC 1051

RESULT 13
ACA24115
ID ACA24115 standard; DNA; 1092 BP.
XX
AC ACA24115;
XX
DT 19-JUN-2003 (first entry)
DE
DE Prokaryotic essential gene #5772.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KM drug design; gene.
XX
OS Borrelia cepacia.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 26-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342922P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JM;
PI Wall D, Trawick JD, Cairt GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX
XX P-PSDB; ABU20245.
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to

PT Isolate candidate molecules for rational drug discovery programs.
PS Claim 14, SEQ ID NO 11985; 1766bp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1092 BP; 175 A; 396 C; 362 G; 159 T; 0 U; 0 Other;
Query Match 20.8%; Score 264.4; DB 8; Length 1092;
Best Local Similarity 56.2%; Pred. No. 6.7e-66;
Matches 540; Conservative 0; Mismatches 411; Indels 9; Gaps 2;
QY 14 GCTTCATACACACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 73
DB 63 GCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
QY 74 GCGAGTGTCTCTACAGTGCATCATCTTGGAGGGCAATGATCATGCCAAATTTGAA 133
DB 123 GCGGCTGCTGCGGACACCGCGCGCAATGCGCGCGCATGCTCATGCGGAACCTGAA 182
QY 134 GCTTCATACACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 193
DB 183 GCGGCGGTCACGACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 242
QY 194 ACCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 253
DB 243 GCGGCGGCGGATGACGTTTCAGCGCGGATGACGTTTCAGCGGACGACGACGCGCGC 302
QY 254 TATGAATATCAATCAGACAG 313
DB 303 CGAGGAATCGCGCGCGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
QY 314 TGTGTCACGACAAATTTCTCAAGTGAAGTACTGATCTTTTTCGGGAAGTTTACAGT 373
DB 363 AGGCGCACGACGAAATTCGACCAATGGGATGACCAATC--TTCGGAATATGCGAAGAC 419
QY 374 TCTCAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 433
DB 420 GCTGAGGCGATGACGAGAAACCGGCAATCGCTGCTGCTGCAACGCGGAGTGAACGATGC 479
QY 434 TGAGTTGACATGTTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 493
DB 480 GTGATGACCTGTTGACCGCGGAGAGAGTGTTCATGACCGCGGATGAGACCGCGCTGCG 539

```
QY 494 GCAGAAATTCACAAATTGAGTGTGATGAGCATGTTACCACTTGAATGCTGTTAA 553
D 540 CCGGAGATTCCTGGGAGCCCTGAAGTGTGTTGAAACAATCAACGAGAGACGGGCGGA 599
QY 554 GTTTGT-----TGAATCTTGACCTGAAGATTTGTTGACGAACCTGTCAACCCACAA 607
D 600 CTAGTCCGCGATGCCGACGGCGGCCCTGCTCGGGCGGACGATCAACCGGCACTCA 659
QY 608 TCTTGTTTGAAACGAATTCCTCTTCCAAAGGGGCTTACACCGCATATTAATGCTCT 667
D 660 CTTCTGTACAAACGCAACGGCTGTTCTGTGGGGGATCCGCCGATTACTATCTCTCT 719
QY 668 TCCAGTCTCTCAAAAGAGATCCACAGAGGACTTGTGTACAGCTGTAAACAATGAG 727
D 720 GCGGTTCTGAAGGCGAGACGCAATCCCTGTGCTGTGAGGCGCGACGCTCGGCA 779
QY 728 TAAAGATTTTCTTGGGACTGATGTCTCTCATGATGACGAAGAAAAGATGTTT 787
D 780 CCGCGCTTCTCTTCCGCGACCGACGCGCGCGCAAGCGCGCAACGGAAGAACGCG 839
QY 788 TTGTGATGTGTGTGATTTTACATGACCTGTACCTTGTGACATATATCGAAGGTGT 847
D 840 GTGGGTTTGGCGGGCTGTCTACACGGCGCTGCAACGCTGATGACCGGAAGGTT 899
QY 848 TGAAGAGAAATGCACTGACCAAGCTTGAAGATTCATAGCTTCAATGACCAATTT 907
D 900 CGACACGGCGGGCGCTGACAAAGCTGGAAGGTTGCGAGCTTCTTGGCGCCGATTT 959
QY 908 TTATGGCTCTCTTGAACCACTCAAAATTAAGTGAATGAACGCGATGGAAGTACC 967
D 960 CTAGGGCTCGCGCGACCGCGACGCTGACGCTGCGCGGAGCGGTGGGAACCTGCC 1019
```

RESULT 14
ABA92787_3
Continuation (4 of 7) of ABA92787 from base 300001 (Buchnera sp. genomic DNA SEQ ID NO:1)
WP Sequence Split Info 7 fragments LOCUS ABA92787 Accession ABA92787

WP	Fragment Name	Begin	End
WP	ABA92787_0	1	110000
WP	ABA92787_1	100001	210000
WP	ABA92787_2	200001	310000
WP	ABA92787_3	300001	410000
WP	ABA92787_4	400001	510000
WP	ABA92787_5	500001	610000
WP	ABA92787_6	600001	640681

Query Match 20.8%; Score 264.2; DB 6; Length 110000;
Best Local Similarity 54.0%; Pred. No. 6; 6e-65;
Matches 564; Conservative 0; Mismatches 478; Indels 3; Gaps 1;

```
QY 20 AATCACAACAACCTGATGATGCGATCTTCTCCGTGATGTGATGTTCTTAAGCAGT 79
D 68078 AATCATTAATAACGTATGATGATGATGATGATGATGATGATGATGATGATGAT 68137
QY 80 TGTCTCTCACAGTCATCACTTTGGAGGGCAATGTCATGCCAAATTTGAAGCTCC 139
D 68138 TATTAAGTATACGTAAATTTTATTAAGAAGCTGTTATTAATGCTCAAACTTAATGTC 68197
QY 140 TATCATCAACACGTGCTGCTGTAGCATACCGGAGGAGATTTGAATCTTAACCTT 199
D 68198 GATTAAGATGTTGTTGAAAAAGCATTCGTATCTTAATGAATTTTAAATCAATGCACTT 68257
QY 200 TGATAGTATTCACACCTCTTATGACATTTATTTGACAGATACCAACGATCTATG 259
D 68258 GAATTAATTAATCAACCGTATTAATGATCTGTTATTTAATCAATTCACAAAGCCATA 68317
QY 260 AATCAACATGCAAGAGAGACGAGTGTATTTGGGTGAAGTTTACCTGCTGTGTC 319
D 68318 ACTAGATTTGTTTCTTAAAAAATTTTGAAGAGCTTAATCTATCCAAATGGTTG 68377
QY 320 CACGACAAATTCACAGATGAGTGAATCTTTTCCGAGAGTGTTCACAGTTCTACA 379
```

```
D 68378 CACGACAAATTCAAAAAAGTGTATTAATAAAAAAT---AGTGAATTAATCTCTGTTTGA 68434
QY 380 AGAATGTTGAGCATATATATGCTCTGCTGTGTTGAGAGAGTTACTTAATCTGAGGT 439
D 68435 ATGTATGAAAAAATAGAAATGCGATTAATTAATTCATGTGTGAAGATTAACAAAAAT 68494
QY 440 TGACATGTTGATGAGAAAGATATTCATTTGAACAGTTCTTAAGACCTTGTGTGAAA 499
D 68495 TGAATCTATGATGAGAAAGCAAAATTTATTAATAAAAAAGTTAATGATCTTTTAAGAAAA 68554
QY 500 ATTTCACATTTGAAGTGTGATGAGCATGTTACACCATTTGATGCTTTAAGTTGT 559
D 68555 ATTTCAAAGTTAAGATATGATTAAGACATTTACACCAAAAGATCTGTGAATAT 68614
QY 560 TGAATCTTGCATGAAAGATTTGTTGACGAACCTGTCAACCCACACATCTTGTGTTGA 619
D 68615 CAATAATACGATGTTAATTAATCTTATCTGCAACTTTACACCAATCAATTTAATCTAAA 68674
QY 620 CAGAAATCTCTCTTCCAAAGGGGCTTACACCGCATTAATTAATGCTTCCAGTCTCA 679
D 68675 TCGAAATGATATGTTTATGATGAGATTCACATTAATCTTATCTTACGATTTTAA 68734
QY 680 AAGAGATCCACAGAGAGCACTTGTGTCACTGTGTAACAGTGAAGTAAAGATTTT 739
D 68735 AAAAAACAAACATCGAATGCACTTAAGAAAAGCCATCTTAATGAGATTAACATTTTT 68794
QY 740 TCTTGGAGTGAATGCTCTCTCATGATGACGAAGAAAGAGTGTCTTGTGATGTGC 799
D 68795 TTTAGGAACGATACAGCTCCACATCTTCAATTAATTAATTAATGCTTGTGATGTGC 68854
QY 800 TGTATTTTACATGACCTGTGACCTGTGATGATGATGAGAGGTTTGAAGAAAGAAA 859
D 68855 GGTATATTTAATGCTCATCATCTCTATTAATCTTATGTTAAGTATTTAGAGAAATG 68914
QY 860 TGCATCTGACAAAGCTTGAAGCATTCATCACTTCAATGAGCAGATTTTATGCGCTTC 919
D 68915 AOCGTTAAATATTTTACATCTTTTGTCTGAAATATGCTTAATTTTATTAACATGCC 68974
QY 920 TGAAGAACACTCAAGATTAAGTGAATGAGCGCATGGAAGTAAACCGAATCTTTTC 979
D 68975 AATTAACAAAGAAACATTAATCAATTAATCAAAAAACGTTAAATTAATTAATAATA 69034
QY 980 TTATCATCAAGAGATTAATTTCCATGTTTCTGTGTAAGTATGCTGATGCTGTTGCCG 1039
D 69035 TGTGGAAGAAATGATTAATTCATTTTATTCAGGTGAATTTTAACTGTTCAATTGA 69094
QY 1040 TCTCTCTGAGAAATATTTGATTT 1064
D 69095 AAGTATTAATAATCAAGATATTT 69119
```

RESULT 15
ACAL9166
ID ACAL9166 standard; DNA, 1047 BP.
XX ACAL9166;
XX
XX 19-JUN-2003 (first entry)
XX
XX Prokaryotic essential gene #823.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX Escherichia coli.
XX
XX W0200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on:      February 12, 2005, 02:16:36 ; Search time 253 Seconds
              (without alignments)
            8220.189 Million cell updates/sec
```

Title: US-10-070-277-1

Sequence: 1 ttgcaaaaatggagctctca.....tgaagattactgtataaaa 1271

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 24055688

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :
Issued Patents NA:*
1: /cgn2.6/prodata/1/ina/5A_COMB_seg.*
2: /cgn2.6/prodata/1/ina/5F_COMB_seg.*
3: /cgn2_6/prodata/1/ina/5A_COMB_seg.*
4: /cgn2.6/prodata/1/ina/6A_COMB_seg.*
5: /cgn2.6/prodata/1/ina/PTCus_COMB_seg.*
6: /cgn2.6/prodata/1/ina/backfile1_seg.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	281	22.1	1071	4	US-09-543-681A-113	Sequence 115, Appl
2	267.4	21.0	1041	4	US-09-328-352-1415	Sequence 113, Appl
3	264.2	20.8	640681	4	US-09-790-988-1	Sequence 11, Appl
4	243	19.1	1068	4	US-09-540-226-410	Sequence 410, Appl
5	242.6	19.1	1197	4	US-09-252-991A-920	Sequence 920, Appl
6	242.6	19.1	2262	4	US-09-252-991A-1041	Sequence 1041, Appl
7	238.2	18.7	92407	4	US-09-596-002-16	Sequence 36, Appl
8	221.6	17.4	1059	4	US-09-489-039A-418	Sequence 418, Appl
9	129.8	10.2	1266	4	US-09-252-991A-1008	Sequence 1008, Appl
10	79.6	6.3	261	4	US-09-252-991A-1117	Sequence 1117, Appl
11	78	6.2	615	4	US-09-252-991A-963	Sequence 963, Appl
12	79	6.1	294	4	US-09-252-991A-1084	Sequence 1084, Appl
13	77.6	6.1	1098	4	US-09-248-708B-4003	Sequence 4003, Appl
14	55.8	4.4	1141	4	US-09-806-708B-22	Sequence 22, Appl
15	51.2	4.0	255	4	US-09-252-991A-962	Sequence 962, Appl
16	38.8	3.1	1141	4	US-09-806-708B-22	Sequence 22, Appl
17	38.6	3.0	36023	4	US-09-949-016-15577	Sequence 15577, Appl
18	38.4	3.0	601	4	US-09-949-016-171450	Sequence 171450, Appl
19	38.4	3.0	126582	4	US-09-949-016-16597	Sequence 16597, Appl
20	36.8	2.9	103750	4	US-09-949-016-13319	Sequence 13319, Appl
21	36.2	2.8	1468	1	US-08-381-433A-7	Sequence 7, Appl
22	36.2	2.8	1468	1	US-08-381-401-7	Sequence 7, Appl
23	36.2	2.8	1558	4	US-09-799-978-11	Sequence 11, Appl
24	36.2	2.8	1600	4	US-09-799-978-11	Sequence 11, Appl
25	36.2	2.8	2110	4	US-09-016-434-1070	Sequence 1070, Appl
26	36.2	2.8	2110	4	US-09-799-978-9	Sequence 9, Appl
27	36	2.8	601	4	US-09-949-016-167519	Sequence 167519, Appl

ALIGNMENTS

C	28	36	2.8	10774	4	US-09-826-509-446	Sequence 446, App
	29	36	2.8	58133	4	US-09-949-016-16464	Sequence 16464, A
C	30	35.8	2.8	1318	4	US-09-086-438-34	Sequence 34, App1
	31	35.6	2.8	601	4	US-09-949-016-62993	Sequence 62993, A
	32	35.6	2.8	601	4	US-09-949-016-109652	Sequence 109652, A
	33	35.6	2.8	601	4	US-09-949-016-109653	Sequence 109653, A
	34	35.6	2.8	601	4	US-09-949-016-206622	Sequence 206622, A
	35	35.6	2.8	832	4	US-09-621-976-2813	Sequence 2813, App
	36	35.6	2.8	101	4	US-09-976-594-487	Sequence 487, App
	37	35.6	2.8	11318	4	US-09-949-016-13564	Sequence 13564, A
	38	35.6	2.8	11318	4	US-09-949-016-12032	Sequence 12032, A
	39	35.6	2.8	40655	4	US-09-949-016-15919	Sequence 15919, A
	40	35.6	2.8	40655	4	US-09-949-016-16191	Sequence 16191, A
	41	35.6	2.8	78649	4	US-09-949-016-14620	Sequence 14620, A
	42	35.6	2.8	78649	4	US-09-949-016-16227	Sequence 16227, A
	43	35.6	2.8	78649	4	US-09-949-016-16228	Sequence 16228, A
	44	35.6	2.8	86116	4	US-09-949-016-14766	Sequence 14766, A
C	45	35.6	2.8	247239	4	US-09-949-016-17590	Sequence 17590, A

RESULT 1

```

: Sequence 113: Application US/09543681A
: Patent No. 6605709
: GENERAL INFORMATION:
: APPLICANT: GARY BRETON
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
: TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 2709.1002-001
: CURRENT APPLICATION NUMBER: US/09/543,681A
: CURRENT FILING DATE: 2000-04-05
: PRIOR APPLICATION NUMBER: US 60/128,706
: PRIOR FILING DATE: 1999-04-09
: NUMBER OF SEQ ID NOS: 8344
: SEQ ID NO 113
: LENGTH: 1071
: TYPE: DNA
: ORGANISM: Proteus mirabilis
: US-09-543-681A-113

```

Query Match	22.1%	Score 281;	DB 4;	Length 1071;
Best Local Similarity	56.1%	Pred. No. 7.2e-78;		
Matches 571;	Conservative	0;	Mismatches 440;	Indels 6;
				Gaps 2
Qy	CTCTCAATCACAACAACCTGATGATTGGCATCTTCATCTCCGTGATGATGATGTTCTTAAG	74		
Db	43 CTCACTATTGCCCGCTCGTAGATTTGGCATGTTCACTTTCGGTAGATGACATGCTAAAA	102		
Qy	75 GCAGTGTCTCTCAAGTGCACATTCACCTTGGAGGGCAATAGTCATGCCAAATTGAAG	134		
Db	103 ACCGTGTTCTTATACAGTGGTATTTTGGCAGAGCTATCGTAAATCCAAATCTTGTT	162		
Qy	135 CTTCTTATCACTACCACTGCTGCTGCTTAGCATTCGGGAGGCGCATATTGAATCTTAA	194		
Db	163 CCCCAGATCACACCATTTGAAGCGCGCCGTGTTATCGTATCGCATAAAGACGATATT	222		
Qy	195 CCTGTTGATAGTGAATTCGAACCTCTTATGACATTTATTTGACAGATCAACACAGTCT	254		
Db	223 CCTAGTGGCGATTAATTTGAAACCATTAATGACCTGCTATTTAACGGATAGCATCTGGCA	282		
Qy	255 ATGGAATCACTAATCTGCAAGAGAGACAGGCGTATTTGGGGTGAAGTTGTACCTGCT	314		
Db	283 TCAGAAAGTGAACAAAGTTTTTTACAAAGCGATTTACTGCTGTAAACTTATCTGTGC	342		
Qy	315 GATGCGACGACAAATTCCTCAAGATGAGTGAATCTTTTCGGGAAGCTTTACCAAGTT	374		
Db	343 AATGGACGACAAATTTAGCCATGTGTTTCCGATAT---CAATAAATCTAATCTAAT	399		
Qy	375 CTACAGAAATGTTGACATTAATGTGCTCTGCTGATTCATGAGAGGTTACTAATCT	434		

147 ACCACTGCTGCTGCTGTGACATACCGGAGGCGAATATGAATCTTTACCTGTGATAGT 206
148 ACAGTAAAGAAAGCTTTTAGCTTATCCCGAACCGCATCTTGGTCATGTTCCGAAAGGCAT 207
207 GATTCAACCCCTCTTTATGACACTTTATTTGACAGATACAACTAGTCTTATGAAATCAAA 266
208 AATTGTAGCCCTGGTATGATGCTTTATTTTATCTGACCACTTCGACTGATGAAGTTCGT 267
267 CTACAGAGAGAGCCAGGTCGTATTTGGGGGTGAAGTTGACCTCGTGTGTCACAGACA 326
268 AAAATTAAATAATCAACATGTAATCGATTAAAGCTTATCTGCTGGTGGACACCA 327
327 AATTCTCAAGTGAAGTGAAGTCTGATCTTTTGGGGAAGTCTTACCATGTTCTACAGAAATG 386
328 AACTCGATPATGTGTGATGATATTT---GTAAAGTTATGCAATGACATTA 384
387 GTTGAGCATTAATATGCTCTGCTGTGCTTCATGAGAGAGTTATCTATCTGAGGTGACATG 446
385 GAAGACATCAAGTTCCTGTATATGCTTCATGTTGTAATCACTCATATCATTTAATATTT 444
447 TTGTATGAGAAAGATATTATTTAGTAAACGGTCTTAAGACCGTGGTGCAGAAATTTCCA 506
445 TTGTATGCGTAAAGCGTCTCTGATGAGATATTTACCTCTATATGAAACAGTTCGCG 504
507 CAATTGAAGTGTGATGAGACATGTTACCAACATGATGCTGTTAAAGTTGTGTAATCT 566
505 AAATCTTAAAGTGTGCTTGAGACATCATACACAGATGAGACACATTTGTT---TTA 561
567 TGCATGAAGATTTGTGACAGCAACTGTACCCCAACAATCTTTGTTGAACAGAAAT 626
562 GAACAAGACCGTAATGTGGCGGACAAATCATCTCCACAACATTTATTTAAACCGTAAT 621
627 TCTCTTCCAGGGGGCTTACAACCGCAATAATTAAGTCTTCAGTCTCTCAAAAGAGAG 666
622 GATATGTTGTGGGTGATTTAAACCGCATCTTCACTGTTTACCAATTTTAAACGCCAA 661
667 ATCCACAGAGGAGCACTTGTGTGAGCTTGAACAATGGAACATAAAAGATTTTCTTGGG 746
662 ACGCATCAAAACGACTTTATCTGAAGTGAAGCAACGCCGTAAATCTTAAATTTTCTTAGGT 741
747 ACTGATGTGCTCTCATGATATAGAGAAAGAAAGAGTTCTTGTGATGTGCGTAAT 806
742 ACAACACGCCCCCTATGACCAAAATGCAAAAGAGATGCTTGTGATGTGACGCTGC 801
807 TACAATGACACTGTAGCCTTGTCAATATATGCAAGGTGTGAAAAGAGAAATGCACTC 866
802 TATGTGCACCAATATGACATGAGCTTTATATGACAAACATTTTACCAAGTCGGTAATTA 861
867 GACAAGTTGAAGCACTCACTAGCTTCAATGAGACAGATTTTATGAGGCTCTCTAGAAAC 926
862 GAGCGCTTGAAGGTTTGTCTAGCCATTTTGTGTGACAGACTTCAAGGCTTACACAGTAAT 921
927 AACTCAAGATTAAAGTTAGTAAAGCAGCGCATGGAAGGTGACCGAATCTTTTCTTAT 983
922 ACTTCTACCATTAATCTTGTGTTAAAGAAATTAACCTGTTCCAGAAATCTTTGATTAAT 978

SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 20.8%; Score 264.2; DB 4; Length 640681;
Best Local Similarity 54.0%; Pred. No. 6,6e-71;
Matches 564; Conservative 0; Mismatches 478; Indels 3; Gaps 1;

```
QY 20 AATCAACAACCGATGATGGCATCTTCATCCCGATGATGATGTTCTTAAAGCAGT 79
    |||
DB 368078 AATCATTTAAACCTGATGATGGCATGTTCTTAAAGACAAATGAGATTTTAAATCAAGT 368137
    |||

QY 80 TGTCTCTCAGACGTCATCATCTTGGAGGCAATAGTCATCCAAATTTGAAGCTCC 139
    |||
DB 368138 TATTAAGTATCTGTTAAATTTTATTAAGAGCTGTTATATGCCAAATCTTATATGTC 368197
    |||

QY 140 TATCACTACCACTGCTGCTGCTGTAGCATACCGGAGGCGATATGAAATCTTACCTG 199
    |||
DB 368198 GATTAAGAGTGTGTTGAAAAGCATGCGTATCGTAAATGAAATTTTAAATCAATGCACCT 368257
    |||

QY 200 TGAATGATGATTTCAACCCCTTTATGACCTTATTTGACAGATACACGATCTTANGA 259
    |||
DB 368258 GAATTTTAAATTTCAACCGTTAAAGCTGTGTTATTTAACTTAATTCACAGCCCTAAAGA 368317
    |||

QY 260 AATCAACTAGCAAGAGAGAGCCAGTCTATTTGGGGTAAAGTTGATCCCTGCTGTC 319
    |||
DB 368318 ACTAGATTTGTTTTTTCTTAAATAATTTGTAAGAGCTAAATTTCTATCAATGTTG 368377
    |||

QY 320 CACGACAAATTTCAAGATGAGTACTGATCTTTTGGGAGTGTTTACAGTTCTACA 379
    |||
DB 368378 CACGACAAATTTCAAACTGTTAAATAAAATTT---AGTATATTCTCCTGTTTGA 368434
    |||

QY 380 AGAAATGTTGAGATTAATTTGCTGCTGCTGCTCATGAGAGGTTCTAATCCTGAGT 439
    |||
DB 368435 ATGTATGAAAAAATAGAAATGCAATTAATTAATGATGTAAGATTAACCAAAATAT 368494
    |||

QY 440 TGACATGTTTGATGAGAAAAAGTATTCATTGAACGGTTCTAAGACCGTTGTGTCAGAA 499
    |||
DB 368495 TGATATCTATGATGAGAGAAATTTATTTGAAAAAAGCTTATGATCTTTTAAAGAAAAA 368554
    |||

QY 500 ATTTCCAAATTTGAGGTCGTGATGAGCATGTTACCAACCTTATGCTGTTAAGTTGT 559
    |||
DB 368555 ATTTCCAAATTTAAGATAGTATTAAGAAATATTAACCAAGAAATCTGTAAGATTAAT 368614
    |||

QY 560 TGAATCTTGACAGTGAAGATTTGTTGACAGCACTGTCAACCCACAAATCTGTTTGA 619
    |||
DB 368615 CAAAAAATTAACGATGTTAATTAATCTTAATCTGCAACTAATTAACACCAATCAATTAATGCTAAA 368674
    |||

QY 620 CAGAAATTTCTCTTCCAGAGGGGCTTAACACCGCATTAATTAAGCTTCAGTCTCAA 679
    |||
DB 368675 TCGAAATGATATGTTTATGTTGAGGATTAACCATATCTTTACGTTTACGATTTTAA 368734
    |||

QY 680 AAGAGATTCACAGGAGGAGCACTTGTGTCACTGTAAACAAGTGAAGTAAAGATTTT 739
    |||
DB 368735 AAAAAAACAACATGAGATGAGCACTTAAGAAAAAGCCATCTTAATGAGATTAACATTTT 368794
    |||

QY 740 TCTTGGAGCTGATAGTGTCTCTCATGATAGACGAAAGAAAGAGTCTTGTGATGTC 799
    |||
DB 368795 TTTTGGAAAGCATTAAGCTTCAACATCTTCAATTAATTAATTAATTAATGCTGGAATGTC 368854
    |||

QY 800 TGTATTTTACAAATGACCTGTAGCTTGTCAATATATGCAAGGTTTGAAGAAAGAA 859
    |||
DB 368855 GGGTATATTTTATGCTCATATCTATTAATCTTTAATGTAAGATTTGAGAAATGAG 368914
    |||

QY 860 TGCACTGACAGCTTGAAGATTCATAGCTTCATGACCAAGATTTTAAAGGCTTCC 919
    |||
DB 368915 AGCGTTAAATATTTAATCAATCTTTTGTCTGAATAATGCTCTAAATTTTAAATGATGCC 368974
    |||

QY 920 TAGGAAACAATCAAGATTTAAGTTAGTAAGACCGCATGAAGGTAACCGGATCCTTTTC 979
    |||
```

```
DB 368975 AATTAACAAGAAACCATTAACATTAATCAAAAAACCGTGTAAATTTTAAAAATAAA 369034
    |||
QY 980 TTATGATGAGAGATTAATTTCCATGTTTGTGTGTAATGCTGAGTGTGCGGC 1039
    |||
DB 369035 TGTGGAAAGAAATGATTAATTTCCATTTTATATCAGGTGAATTTTAACTGTCAATTGA 369094
    |||

QY 1040 TCTCTCTGAGATCATTTGTCATT 1064
    |||
DB 369095 AAGTATTAATAATCAAGATATT 369119
    |||
```

RESULT 4
US-09-540-236-410

```
; Sequence 410, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATA.
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 410
; LENGTH: 1068
; TYPE: DNA
; ORGANISM: M. catarrhalis
US-09-540-236-410
```

Query Match 19.1%; Score 243; DB 4; Length 1068;
Best Local Similarity 55.3%; Pred. No. 7e-66;
Matches 552; Conservative 0; Mismatches 420; Indels 27; Gaps 3;

```
QY 6 AAAATGAGCTCTCAATACACAACTGATTTGGCATCTTATCTCCGTGATGATGAT 65
    |||
DB 7 ACAAAGCATTAATCTCATTCAGCAGAGATCTGCGATTCATCTGCGATGATGAA 66
    |||

QY 66 GTTCTTAAGCAATGTCCTCAAGTCATCACTTTGGAGGCAATAGTCATGCCA 125
    |||
DB 67 GCTCTGCAACCAACCGTTCGAGTGGGCAACCAATTTAAACGTGATTTGATGCT 126
    |||

QY 126 AATTTGAAGCTCTTATCACTACCATGCTGCTGTGATACATCCGAGGCGATATG 185
    |||
DB 127 AATTTGCTGCACTGTTAAATCAATCAAGCCATGCTTATGCTGAACGATCTTA 186
    |||

QY 186 AAATCTTAC-----CTGTGATAGTATTTCAACCTCTTATG 224
    |||
DB 187 AAGCATTAAGAAGACATGATTTAACAGCCATCGCAAAAGATGTTGATCCAGCATG 246
    |||

QY 225 ACATTTATTTGACAGATTAACCAACGCTTATGAAATCAAACTAGAGAGAGCGAG 284
    |||
DB 247 GTGCTTTATTTGACCGATCAAAACCGCAAGATATGATGATGCGCACAAACAGT 306
    |||

QY 285 GTGATTTGGGGTGAAGTGTACCTGCTGTGTCGACGAAATTTCTCAAGATGAGTG 344
    |||
DB 307 ATGTATGCGCAATTAAGCTGTATCTGCTGTGAGCAACCAATTTACGATGATGTA 366
    |||

QY 345 ACTGATCTTTTTCGGAAGTGTTAACAGTTCTACAGAAATGTTGAGCATATATGCT 404
    |||
DB 367 ACTGATATTTTTCGACCAAGCTT---GGTATTTAGCCATGCAAAAGCATGCGATCCG 423
    |||

QY 405 CTGCTGTGATGAGAGGTTTACATATCTGAGGTTACATGTTGATGAGAAAGGTA 464
    |||
DB 424 CTACTGTGTGATGAGGTTAATTAACGATGACAACTATTAATCTTTGATCTGTAAGAAAGCA 483
    |||

QY 465 TTCAATGAACGTTCTTAAGACCGTGTGTCGAAATTTTCCAAATTTGAAGAGTGTG 524
    |||
DB 484 TTTTGTGATGATGCTTAATTAATAATATACAAATTTCCGAATTTAAATAATGTGATG 543
    |||

QY 525 GAGCATTTAACCAATGATGCTGTTAAGTTTGTGATTTTGCACATGAAGATTTGT 584
    |||
DB 544 GAGCATATCACTACCGCGATGACGAGATTTTGTCTTGTCT---CAAGCAATCAATC 600
    |||
```

QY 585 GCAGCAACTGTACCCCAACAATCTGTTTAAACGAGATTTCTCTTCCAAAGGGGC 644
 Db |||||
 QY 601 GCAGGACAATTACCCCGACAGATCTTTATTTAACGAAATCATCTCTGTTGGCGGT 660
 QY 645 TTACAACGCGCATTAATTAAGTCTGCCATCTCCAGTCTCAAAAAGAGATCCACAGGAGGACATT 704
 QY 661 ATTAAGCCGCACTATTATGTTGGCCATTTAAACCGTGCGACACCAAAACGTTTG 720
 QY 705 GTGTACGCTGAACAGATGAGATGAATAAGATTTTTTCTTGGGACATGATAGTCTCTCAT 764
 Db 721 CTGAAGGCGCAACAGTGCATCCTAAGTTTTTCTTAGGACAGATTCGCCCCCAT 780
 QY 765 GATGAGAGAGAAAAGAGTCTTTGTGATGTGCTGTATTTCAATGCACCTGACGCC 824
 Db 781 GCCACAACAAGAAAGATCATCGTGTGCTGTGACAGGTGTATTAAGTCTTGACAGCT 840
 QY 825 TTGTCACTAATATCGAAGGTGTTGAAAAGAAAATGCACTGCGCAAGCTTGAAGATTC 884
 Db 841 CTACCACTTAAGCCATGCGATTGAAATGTGCAATGCACTGTATTAACCTGAAAACTTT 900
 QY 885 ACTAAGCTCAATGACCAAGATTTTATATGCGCTCCTAGAACAACATCAAAAGATTAGTGG 944
 Db 901 GCCAGCGCATATGATGCACAATTTTATATGCTTAGCAATCAACACCACTCAAAATCACCCTA 960
 QY 945 AGTAAGCGCATGGAAGTACCCGAATCTTTTCTTAT 983
 Db 961 ATTAACCAACACAGCAACATTAAGTTGAGATTTTATAT 999

```

RESULT 5
US-09-252-991A-920
; Sequence 920, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107156.156
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 920
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-920

```

Query Match	19.1%;	Score 242.6;	DB 4;	Length 1197;
Best Local Similarity	53.8%;	Pred. No. 1e-65;		
Matches 547;	Conservative	0;	Mismatches 464;	Indels 6;
			Gaps	2
Qy	15	CTCTCAATACCAACCTGATATTGGCATTTTCATCTTCGGATGGTGAATGTTCTTAAG	74	
Db	163	CTCACCTCTCTGGCGCCCGACACATGGCACTTCACTCGCGAGACGGTGGCGCGCTGGCC	222	
Qy	75	GCAAGTTGTCTTCAACAGTGACATCACTTTGGAGAGGCAATGTATATCCAAATTGAG	134	
Db	223	AATACCGTCGGCGAGCGCGCCGCACTTTGGCGCGCGCATGTATATCCAAACTGGTG	282	
Qy	135	CTCTCTATACACACATGCTGTCTGTAGCATCCGGAGAGGCATATTGAAATCTTTA	194	
Db	283	CGCGCGGTGGCGACGCGCGGAACCGACGCCATACCGGAGGCCATCTCTGCGCGCGCG	342	
Qy	195	CTGTGTAGTAGATTTCACCCCTCTTATGACATTTATTTGACAGATATACCAACAGTCT	254	
Db	343	CCCGCGCGCAACCGCTTGAAGCCCGTATGTCGTACCTCACACGACCGCAGGCGCC	402	
Qy	255	ATGAAATCAAACTAGCAAGAGAGCCAGGTGATTTTGGGGTGAATTGTACCTCTGT	314	

Db 403 GAGGAATCCGATCCGCCAAGGCCCTTCGTGATGCGCCAACTCTATTCGGCA 462

QY 315 GGTGCACACAAATTTCTCAGAATGAGGATCTATTTCCGGAAGTGTTCAGATT 374

Db 463 GCGCCACACCAACTCGGATTCCTGCGTAC--CCGATCGACAAATTTCCGAACG 519

QY 375 CTACAAAGAAATGGTTGACATATATGCTCTGCTGTGATCAGAGAGTTACTAATCT 434

Db 520 CTCGAGCGATGCGGAGTCCGATGCGCTCGTGTGATGCGAGGTGACCCCGGCC 579

QY 435 GAGGTGACATGTTGATAGAAAAGGTATTCATTGAAACGGTTAAAGCCGTGGTG 494

Db 580 GAGGTGCAGATTCATCGATGAGAAAGATTCATCAGACGACCTGCGCGGTGGTC 639

QY 495 CAGAAATTTCCAAATTTGAAGTGTGTATGAGCATTTACACCATTTGATGCTTTAAG 554

Db 640 GAGCGCTTCCGACCTTGAAAGTGTTCGAGACATTCACACCGCGCAGCGCGCCAG 699

QY 555 TTTGTTGAATCTTGCACGTGAAGATTTGTTGCAGCACTGCACCCCAACATCTTGT 614

Db 700 TTCTTCGCGGAAGCCC--GGCAAGTGGCGGCACATTAACCGCCATCACTGCTG 756

QY 615 TTGAACGGAATTTCTCTTCCAAAGGGGCTTCAACCGCATATTATGCTCTTCAGTC 674

Db 757 TACACCGCAACCAATGCTGTGTCGGCGGATCGTCCGACTTATTGCTCGCGGATC 816

QY 675 CTCAAAAGAGATTCACAGGAGGACCTGTGTCAGCTGTAAACAAGTAAAGTAAGA 734

Db 817 CTCAAGGCAACACCCACGAAAGCCGTGTGAGCCGCGGTGACGCAATCCGAAA 876

QY 735 TTTTTCCTGGAGCTATAGTGTCTCTCATATAGACGAAGAAAAGTGTCTTGTTGA 794

Db 877 TTCTTCTCGGACCGCACTCGGCGCGCACGCCCGCAAGCAAGGAAGCCCGTCGGGT 936

QY 795 TGTGCTGTATTTACATGCACTGTAGCTGTGCTGTAGTATATGCGAAGGTGTAAAG 854

Db 937 TGGCGCGGTGTCTACAGCGCTTACGCGGCATCGAGCTGTATGCCGAGGCTTCGAAACG 996

QY 855 GAAATGACATCGAAGAAGTTGAAGATTCATCTACTTCAATGACCAAGATTTTATGGG 914

Db 997 CGAATGCGCTGACAGCTGGAAGGCTTCGCAAGCTGCAAGCGCCCGACTTCTATGGC 1056

QY 915 CTTCCTAGGAACACTCAAAGATTAAGTTAGTGAAGACGCACTGAAGATCCCGAATCC 974

Db 1057 CTGCGCGCAACACGACCGGATCACCTGTGTCGCGAGATGGCAGGCGCCCGCACGC 1116

QY 975 TTTTCTATGATCAGAGATTTATATCCCATGTTTGTGTGAATGCTCTGACTGG 1031

Db 1117 CTCCTTCGCGAATTCGACGTGTGCGCTGTGGCGCGCGCGAGAACATTCGCTGG 1173

```

RESULT 6
US-09-252-991A-1041/c
; Sequence 1041, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 3142
; SEQ ID NO 1041
; LENGTH: 2262
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1041

```



```

RESULT 8
US-09-489-039A-418
: Sequence 418 Application US/09489039A
: Patent No. 6610836
: GENERAL INFORMATION:
: APPLICANT: Gary Breton et. al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
: FILE REFERENCE: PNEUMONICAE FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 2709.2004001
: CURRENT APPLICATION NUMBER: US/09/489, 039A
: CURRENT FILING DATE: 2000-01-27
: PRIOR APPLICATION NUMBER: US 60/117, 747
: PRIOR FILING DATE: 1999-01-29
: NUMBER OF SEQ ID NOS: 14342
: SEQ ID NO 418
: LENGTH: 1059
: TYPE: DNA
: ORGANISM: Klebsiella pneumoniae
: US-09-489-039A-418

```

```

QY      380 AATAATGGTTGAGCTAATATATGCTTCGTGCTGTTCAATGAGAGGTTACTAATCTGAGGT   439
Db      396 GCGGATGAGAAAAAATCGGGGATGCGGTTGCTGTTCAAGGGAGATCACTCAGCCGAGAT   455
QY      440 TGACATGTTTATATGAGAAAAAGTATTCATTGAAACGGTTCTAAGACCGTTGGTGCAA   499
Db      456 TGAATATTTTCATGCTGAGAGCGGCTTTATTTGAGACGATATGAGAGCCCTCGGCAGCG   515
QY      500 ATTTCACAAATTGAAGGTCGTGATGAGACATGTTACCAACCATGATGCTGTAAAGTTGT   559
Db      516 CCTGCCCGGCTGAAAGTGTGTTTTTGAACATATCACACCAAGAAAGCCCGCAGATACGT   575
QY      560 TGAATCTTGCACTAAGAGATTGTTGAGACATGTCAACCCCAACATCTGTTTGA   619
Db      576 ---GCGTGAAGGTAATAATCACTGCTGGCTGTACATACACCCCGACATCTGATGTTCAA   632
QY      620 CAGAAATTTCTCTCTTCCAGAGGGGCTTACACCCGACATTAATTACTGCTCCAGTCTGCA   679
Db      633 CCGGACCAATATGCTGTGTGGGCGGCAATGTGCGCTTACATGTAATCTGTCTGCCGTACTTA   692
QY      680 AAGAGATCCACAGGAGGCACTTGTGTACGTGTAAACAATGGAAGTAAAGATTTT   739
Db      693 ACGAATATTCATCAGAGGCTCTGCGGAGCTGTGTGCGACCGCTTCAGCGGCTT   752
QY      740 TCTTGGACTGATATGTCCTCATATGATAGCAAGAAAAAGAGTGTCTTGTGGAATGTC   799
Db      753 CCTGGCACCGACTCCGACCGCACGCGCCGACCGTAAGAGACAGCTGTGCTGCGC   812
QY      800 TGGTATTTACATGACCTGTAGCCTTGTGCAATATATGCAAGGTGTTGAAAAAGAAA   859
Db      813 CGGCTCTTTAAAGCCCGACCGCCCTCGGACGATACCCACGGGTGTTGAAAGAAATGA   872
QY      860 TGCACTGCAAGACTTGAAGCATTCATGACTTCAATGAGACAGATTTTATGGGCTCC   919
Db      873 CGCTTGACGACATTGTGAGGCTTTCTGCTCGCTAAACGCGCCGCTTCTACGGTCTGC   932

RESULT 9
US-09-252-991A-1008
; Sequence 1008 Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR APPLICATION NUMBER: US 60/074.788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094.190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1008
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1008

```

Query Match	10.2%	Score 129.8;	DB 4;	Length 1266;
Best Local Similarity	52.9%	Pred. No. 4e-30;		
Matches 303;	Conservative	0;	Mismatches 267;	Indels 3; Gaps 1;
QY	459	AAGGATTCATTGTAAGAAGGTTCTAAGACCGGTGTGTCAGAAATTTCACAAATTGAAGGTC	518	
Db	2	AAGCAGTTTATTCACACAGCAGCACTCGCGCGGGGTGTGAGCGCTTCCCAACCTCGAAGGTG	61	
QY	519	GTGATGAGCAGATGTTTACCAACCAATTGATGCTGTTAAGTTGTGTAATCTTGACATGAAGA	578	
Db	62	GTCTTGACACATCAACCAACGGGCAAGCGCGCCAGTTCTGTCCGGGAAAGCCCC--GGCC	118	
QY	579	TTTGTTCAGCAACGTGTACCCCAACAATCTTGTTTTGAACAGAAATTCTCTCTTCGA	638	

Db 119 AACGTGGCGCGACATTACCGCCATCCTGCTGTACACCGCAACCATGCTGTC 178
Qy 639 GGGGGCTTACAAACCGCATTAATTACTGCTTCACCTCTCAAAAAGAGATCCACAGGAG 698
Db 179 GGGGATATCCGTCCGCACTTCTATTGCTTCCGATCTCAAGGCAACCCACAGGAA 238
Qy 699 GCACTTGTGTACGTGTAAACAGTGAAGTAAAGATTTTCTTGGGATGTATAGTCT 758
Db 239 GCCCTGTGTAGACCGCGGTGTAGAGCGGCAATCCGAAATCTTCTCCGACCGACTCGCG 298
Qy 759 CCTCATATAGAGAGAAAGAGTGTCTTGTGATGTGCTGTATTAATGACACT 818
Db 299 CCGGACGCGCGCAACGCAAGAGCGCCCTGCGTTCCGCGGTCTACAGCGCTAC 358
Qy 819 GTAGCTTGTCAATATATGCGAAGGTGTGAAAGAAAGAAATGCACTGCAAGCTTGA 878
Db 359 GCGGCAATGAGCTGTATGCGAGGCTTGAACAGCGCAATGCGCTGGAAGCTGGA 418
Qy 879 GCATTCACTAGCTTCAATGAGACGATTTTATGCGCTTCTGAGAACTCAAGAT 938
Db 419 GCGTTCGCGACCTGACGCGCGGCTTCTATGCGCTGCGCAACCGAGATC 478
Qy 939 AAGTGTAGTAAAGACGCAATGAGTACCGAATCCTTTCTTATGATCAGAGATAT 998
Db 479 ACCCTGTGTCCGCGAAGATGCGCGCGCCGACCTCCCTTCCGCGACTTGCAGCTG 538
Qy 999 ATTCCATGTTTCTGTGGAATGCTGCACTG 1031
Db 539 GTGCGCTGCGCGCGCGAGACACTCCGCTG 571

RESULT 10
US-09-252-991A-1117/c
; Sequence 1117, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1117
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1117

Query Match 6.3%; Score 79.6; DB 4; Length 261;
Best Local Similarity 57.1%; Pred. No. 1.1e-14;
Matches 145; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 676 TCAAAAGAGATCCACAGGAGGACTTGTCTACGCTGTAAACAGTGAAGTAAAGAT 735
Db 261 TCAAGCGCAACCCACCAAGAGAGCCGTGTGACGCGCGGTGCGCAATCCGAAAT 202
Qy 736 TTTTCTGGGACTGATAGTGTCTCTCATGATAGACGAAGAAAGATGTTCTTGTGAT 795
Db 201 TCTTCTGGGACGATCTGCGCGCGGACGCGCGGCAAGCAAGACCGCGCTGCGTT 142
Qy 796 GTGCTGTATTATTAATGACACCTGTAGCCTTGTCAATATGCGAAGGTGTTGAAAG 855
Db 141 GCGCGCGGTGTACAGCGCGCTACGCGCATGAGCTGTATGCTGAGGCTTGAACAGC 82
Qy 856 AAAATGCACTGACAAAGCTTGAAGATTTCACTAGTTCATATGAGACCAATTTTATGGG 915
Db 81 GCAATGCGCTGACAAAGCTGGAAGGCTTGCAGGCTGTGACGCGCGGACTTCTATGCGC 22

Qy 916 TTCTTAGAACAAC 929
Db 21 TGCGCGCAACACC 8

RESULT 11
US-09-252-991A-963
; Sequence 963, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 963
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-963

Query Match 6.2%; Score 79; DB 4; Length 615;
Best Local Similarity 53.4%; Pred. No. 2.7e-14;
Matches 166; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

Qy 721 GTGAACTAAAGATTTTCTTGTGGAATGATAGTGTCTCTCATGATAGAGAAAG 780
Db 1 GCGGCAATCCGAAATTTCTCTCGGACCGACTGCGCGCGACGCCGCGCAAG 60
Qy 781 AGTGTCTTGTGATGCTGTATTTTACATGACCTGTAGCCTTCTCATATATGCA 840
Db 61 AAGCGCTCGGTGGGTGGGTGTACAGCGCTTACGCGCGCATGAGCTGTATGCC 120
Qy 841 AGGTGTTGAAAGAAATGCACTCGACAGCTTGAAGATTCATAGCTTCAATGAC 900
Db 121 AGGCTTCAAGACCGCAATGCGTGTGACAAAGCTGAGAGGCTTGCAGCGCGC 180
Qy 901 CAGATTTTATGCGCTTCTTGAACCACTCAAAAGATTAAGTATGATAGACCGCAT 960
Db 181 CGACTTCTATGCTCGCGGCAACCGACCGATCACCTGTGCGCGAGGAATGGC 240
Qy 961 AGTACCGCAATCCTTTCTTATGATCAGAGATATATCCATGTTTCTGTGAAA 1020
Db 241 AGGCGCGCGCGAGCTCCCTTGGGCACTTGCAGTGTGCGCTGCGCGCGGAGA 300
Qy 1021 TGCTGCACTGG 1031
Db 301 CACTCCGCTG 311

RESULT 12
US-09-252-991A-1084/c
; Sequence 1084, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW nucleic - nucleic search, using SW model

Run on: February 12, 2005, 05:10:23 ; Search time 811 Seconds
(Without alignments)
9247.207 Million cell updates/sec

Title: US-10-070-277-1

Sequence: 1 ttgcgaataatgagctctca.....tgaagattcactgataaaaa 1271

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5378673 seqs, 2950229984 residues

Total number of hits satisfying chosen parameters: 10757346

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptoddata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptoddata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptoddata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptoddata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptoddata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptoddata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptoddata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptoddata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptoddata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptoddata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptoddata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptoddata/1/pubpna/US10A_PUBCOMB.seq:*
13: /cgn2_6/ptoddata/1/pubpna/US10B_PUBCOMB.seq:*
14: /cgn2_6/ptoddata/1/pubpna/US10C_PUBCOMB.seq:*
15: /cgn2_6/ptoddata/1/pubpna/US10D_PUBCOMB.seq:*
16: /cgn2_6/ptoddata/1/pubpna/US10E_PUBCOMB.seq:*
17: /cgn2_6/ptoddata/1/pubpna/US10F_PUBCOMB.seq:*
18: /cgn2_6/ptoddata/1/pubpna/US10G_PUBCOMB.seq:*
19: /cgn2_6/ptoddata/1/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptoddata/1/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptoddata/1/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptoddata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	646	50.8	1743	18	US-10-451-554-2
2	607.4	47.8	1826	17	US-10-424-599-6649
3	549.2	43.2	1545	18	US-10-437-963-86830
4	531.6	41.8	1893	18	US-10-435-115-86204
5	345	27.1	1258	18	US-10-425-115-86207
6	311.8	24.5	931	18	US-10-767-701-2040
7	281	22.1	1053	17	US-10-282-122A-41490
8	277.2	21.8	1047	17	US-10-282-122A-42220
9	267.4	21.0	1032	17	US-10-282-122A-8516
10	265.2	20.9	1056	17	US-10-282-122A-14242
11	264.4	20.8	1092	17	US-10-282-122A-11985

12	264.2	20.8	640681	9	US-09-790-988-1	Sequence 1, Appl1
13	245.6	19.3	1047	17	US-10-282-122A-7036	Sequence 7036, Ap
14	241	19.0	1047	17	US-10-282-122A-30305	Sequence 30305, A
15	238.2	18.7	1062	17	US-10-282-122A-15202	Sequence 15202, A
16	238.2	18.7	1062	17	US-10-282-122A-26930	Sequence 26930, A
17	238.2	18.7	92407	17	US-10-672-787-36	Sequence 36, Appl
18	236.2	18.6	1044	17	US-10-282-122A-31411	Sequence 31411, A
19	236	18.6	1032	17	US-10-282-122A-29054	Sequence 29054, A
20	231.2	18.2	1035	17	US-10-282-122A-33544	Sequence 23544, A
21	221.6	17.4	1041	17	US-10-282-122A-33743	Sequence 33743, A
22	219.2	17.2	1047	17	US-10-282-122A-39694	Sequence 39694, A
23	218.4	17.2	1047	17	US-10-282-122A-36568	Sequence 36568, A
24	186.4	14.7	1052	17	US-10-282-122A-36568	Sequence 1199, Ap
25	183.8	14.5	642	18	US-10-021-323-1199	Sequence 86208, A
26	178.4	14.0	828	18	US-10-425-115-86208	Sequence 4146, Ap
27	177	13.9	381	11	US-09-732-627A-4146	Sequence 6648, Ap
28	166.6	13.1	434	17	US-10-424-599-6648	Sequence 19356, A
29	149	11.7	570	17	US-10-021-323-2921	Sequence 2921, Ap
30	138.8	10.9	574	18	US-10-021-323-2921	Sequence 107341, A
31	130.8	10.3	411	17	US-10-424-599-107341	Sequence 6104, Ap
32	129.6	10.2	638	18	US-10-282-122A-18083	Sequence 18083, A
33	114.4	9.0	1008	17	US-10-282-122A-32732	Sequence 32732, A
34	105	8.3	407	17	US-10-282-122A-22560	Sequence 22560, A
35	77.6	6.1	1020	17	US-10-282-122A-2523	Sequence 2523, Ap
36	67.4	5.3	348	17	US-10-335-977-2162	Sequence 2162, Ap
37	65.2	5.1	1020	17	US-10-335-977-2162	Sequence 2162, Ap
38	65.2	5.1	1113	17	US-10-425-115-89	Sequence 89, Appl
39	58.4	4.6	2580	18	US-10-425-115-89	Sequence 2160, Ap
40	50.4	4.0	438	17	US-10-425-115-164730	Sequence 164730, A
41	48.2	3.8	1992	18	US-10-027-632-250146	Sequence 250146, A
42	42	3.3	652	13	US-10-027-632-250147	Sequence 250147, A
43	42	3.3	652	17	US-10-027-632-250146	Sequence 250146, A
44	42	3.3	652	17	US-10-027-632-250147	Sequence 250147, A
45	42	3.3	652	17	US-10-027-632-250147	Sequence 250147, A

ALIGNMENTS

RESULT 1
US-10-451-554-2
Sequence 2, Application US/10451554
Publication No. US20040111769A1
GENERAL INFORMATION:
APPLICANT: Rodolphe Arthur Kanhonou
APPLICANT: Ramon Serrano Salom
TITLE OF INVENTION: Sugar beet genes involved in stress tolerance
FILE REFERENCE: 1187-24
CURRENT APPLICATION NUMBER: US/10/451.554
CURRENT FILING DATE: 2003-06-20
PRIOR APPLICATION NUMBER: PCT/EP01/15093
PRIOR FILING DATE: 2001-12-20
PRIOR APPLICATION NUMBER: EP 00870319.1
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 60/271.656
PRIOR FILING DATE: 2001-02-26
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 1743
TYPE: DNA
ORGANISM: Beta vulgaris
US-10-451-554-2

Query Match 50.8%; Score 646; DB 18; Length 1743;
Best Local Similarity 76.4%; Pred. No. 5.2e-175;
Matches 793; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

Qy 3 GCAAAATGAGCTCTCATACACAACTGATGATGGCATCTTCATCTCCGATGCT 62
DB 193 GTAAAGATGAACTGACTCTTACACGCGCTGATGATGCGATTCATCTCCGATGGA 252

Db	1019	TGTCCTCCAGTCGCTCAAAAAGAGATCCATAGACAGGCTATTGTTCCGCTGTCACTAGT	1078
Qy	723	GGAGTAATAAGATTTTTCTTGCGACTGATATGTGCTCTCATGATAGACGAAGAAAAAG	782
Db	1079	GGAGTAATAAGATTTTTCTTGCGACTGATATGTGCTCATGATAGAGCGTAATAAGGA	1138
Qy	783	TGTTCTTGATGTGCTGCTGATTTTACATGCACTGTAGGCTTGCTAGTATATGCCAG	842
Db	1139	TGTTCTTGATGTGCTGCTGCGCATTAATACAACTCACCGGTTGCTATATCACTATATGCCAAA	1198
Qy	843	GTGTTTGAAGAAAAATGCACTCGACAGGCTGGAAGATTCATAGCTTCAATAGACCA	902
Db	1199	GTTTTGAAGAGGCTGCTGCACTGTATATAGAGGCTTTTACAAGCTTTTAAGGACCT	1258
Qy	903	GATTTTATAGGCTCTCTTAGAACCAACTCAAGATTAAGTGAAGTAAAGCCATGAG	962
Db	1259	GACTTCATAGGCTCTCCCAAGAACAAAGTCATAAGATTAACTGAGGAAAGCTCCTGGAAA	1318
Qy	963	GTACCCGAATCCTTTTCTTATGCAATCAGAGATATATATCCCATGTTGCTGCTGAATG	1022
Db	1319	GTACCTGATATATTGTCAATTCATTGGAGACATGTTCCCATGTTGCTGCTGAAC	1378
Qy	1023	CTGCACTGATGCGGCGCTCCTCTGA	1049
Db	1379	CTTGATAGGAGCAATGCTTTGTGGA	1405

RESULT 3

```

/ Sequence 86830, Application US/104379663
/ Publication No. US2004012343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovacic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Boukharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21 (53221) B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ NUMBER OF SEQ ID NOS: 2003-05-14
/ SEQ ID NO 86830
/ LENGTH: 1545
/ TYPE: DNA
/ ORGANISM: Oryza sativa
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_85834C.1
US-10-437-963-86830

```

Query Match	43.2%	Score 549.2;	DB 18;	Length 1545;
Best Local Similarity	70.4%	Pred. No. 4e-147;		
Matches 734;	Conservative	0;	Mismatches 308;	Indels 0;
				Gaps 0

Qy	11	GGAGCTCTCAATACACAACTGATGATTTGGCATCTTCACTCTCCGATAGTGATGTTCT	70
Db	153	GGAGGCTCACCATCACCGGCGCGACATGGCACCTCCACCTCCGGGAAGGGAGTCTCT	212
Qy	71	TAAAGGAGTGTGCTCTCACAGTGCATCATCTTTGGGAAGGCAATGATGACCAATTT	130
Db	213	CGGGGCGGCTCTCCCAACGCGCGATGCATTTTGGGAAGGCGATGTGATGCCCACT	272
Qy	131	GAAAGCTCTTATCATCTACCACTGCTGCTGTAGCATACGGGAGCGCATATTGAAATC	190
Db	273	GAAAGCAACGGGTACCAACGACGACAGAGGTTTGGAGTACAGGAGGAAGATCCTCAGGCG	332
Qy	191	TTTACTGTTGATAGTGAATTTCAACCTCTTATGACATTTATTTAGCAGATPACAACG	250
Db	333	GTGCGCGCTGGGAGCACTTTGTGCGCTGATGAGCGCTTCACTACGAGACATATCTAG	392

OY	251	TCTATGGAATCAAACTACGAAAGAGAGACAGGCTGTAATTTGGGGTGAAGTTTGAACC	310
Db	393	CCCGAGAGAGATCAAGCTCGCAAAAAAGAGTGGCGGTTTTGCTGTGAAGTTGAACC	452
OY	311	TGCGGAGCCACGCAAAATTCCTAAGATGAGAGTACTGATCTTTTGGGGAAGTGTTACC	370
Db	453	TTCTGGAGCACTTACCAACTCCAAAGTGTGTCTCATGATATATTTGGGAAGTCTTCC	512
OY	371	AGTTCACAGAAAATGAGTTGAGCAATAATATGCTGTGCTGTTCAATGAGAGGTTACTAA	430
Db	513	TGTGCTTGAGGAATATGCTAGGCGAGAGATGCTTGTGCTTGTTCATGAGAAATGTACAGA	572
OY	431	TCTCGAGTTCACACTGTTTGTATAGAAAAGGTATTCATTGAAAACGGTCTTAAGACCGTT	490
Db	573	TCAGCATGTTGATACCTTTGATGTGTGAAGAGTTTTCATTGAGAAAATATTGGCACACT	632
OY	491	GGTCGAGAAAATTTCCACAATTGAAGTGTGATGAGACATGTTTACCACTTGAATGCTGT	550
Db	633	TGTACAAAGATCGCACAGCTGAAGAAATTTGTTATGGAACATATCACCATATGATGGGT	692
OY	551	TAACTTTGTTGAATCTTTGCACTGAAGGATTTGTTGCAAGCACTGTCAACCCCAACAATCT	610
Db	693	TAACTGTGTAATTCATGCAAAAGAAAGTCAATGTTGCTGCAACGGTGACTCCCAACAATCT	752
OY	611	TGTTTGAACAGGAATTCCTCTCCAAAGGGGGCTTACAAACCGCATATTAATCTGCTTCC	670
Db	753	TCTTCTCAACAGGAATGCGTTATTTTCAGGGTGGCTTGCAACACACAACTATTTGCTTACC	812
OY	671	AGTCCCTAAAAGAGATCCACAGGAGGCACTTGTGTCACTGTATTAACAGTGAAGTAA	730
Db	813	AGTACTGAAAAGAGAACTCATTAACAAGCAATATGTGTCTCGGGTAAACAAGTGGAGTAA	872
OY	731	AAGATTTTTCCTGGGACTGATAGTGCTCCCTCATGATATAGCAAGAAAGAGTGTCTTG	790
Db	873	ACAAATCTTTCGTGTACTGACAGTGTCTCTCATGATATTAACGAGAAAAGAAATGCTCTG	932
OY	791	TGSAATGTGCTGTATTTTACATGCACTGTATGACCTTGTCAATATATSCGAAGGTGTTGA	850
Db	933	TGTTGTGACAGGATATATATAGCGTCCGTGTGCTCTTCTCTTATGCAAAAGTATTCGA	992
OY	851	AAAGGAAAATGCACTCGACAGCTTGAAGCATTCATCAATAGCTTCAATGAGCCAGATTTTAA	910
Db	993	ACAGGCGCGTCTCTTGATATAGCTATGAGACATTTTACAAGCTTCAATATGGCCAGATTTTAA	1052
OY	911	TGGGCTTCCTATGGAACAATCAAAAGATTAACTTATAGTATGAGACGCGCATGGAAGTACCGGA	970
Db	1053	TGGCTTTCAAAGAACACTTCAAAAGATTGTCTTATGAAAAGCGCCTGGAAGATACCTGGA	1112
OY	971	ATCTCTTTTCTTATGATCAGAGAGATATTTATCCATGTTTGTCTGTGAAATGCTGACCTG	1030
Db	1113	TACATATCAATATAGTTTACGAGAGAGATGTGCTATGTTTACTGGCAACACCTTCGATGT	1172
OY	1031	GTTGCCGGCTCTCTCTGAGAA	1052
Db	1173	GCTTTCATCCGATCAACTTGAA	1194

RESULT 4

```

US-10-425-11586204
? Sequence 86204, Application US/10425115
? Publication No. US20040214272A1
? GENERAL INFORMATION:
? APPLICANT: La Rosa, Thomas J.
? APPLICANT: Kovalic, David K.
? APPLICANT: Zhou, Yihua
? APPLICANT: Cao, Yongwei
? TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
? TITLE OF INVENTION: Plants
? FILE REFERENCE: 38-21(5322)B
? CURRENT APPLICATION NUMBER: US/10/425,115
? CURRENT FILING DATE: 2003-04-28
? NUMBER OF SEQ ID NOS: 369326

```


Db 789 TGTACAAAACCTCCGAGCTGAAAATTGTCATGGAACATATTCACAATATGATGACAT 848
Qy 551 TAACTTGTGTAATCTTGGACACTGAGGATTTGTTGGACAACTGTGACCCACACATCT 610
Db 849 GAACCTCATGATGATCATGTGAAGAAGTCAATGTTGGCAAGAGTGAATCCAGCATCT 908
Qy 611 TGTTTGAACAGAAATTTCTCTTCCAGAGGGGCTTACACCGCATATATTCGCTTCC 670
Db 909 CCTCTCAATAGAAATGCTTATTTTCAGGGTGGCTTGCACCCACACATATTTGCTGCC 968
Qy 671 AGTCCCAAAAGAGATTCACAGGAGGACCTTGTGCTGATGTAACAAGT 723
Db 969 AGTCTGAAAAGAGACTCATAGTATGTCGATGCTCTCTATTAAGAAATG 1021

RESULT 6
US-10-767-701-2040
; Sequence 2040, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767, 701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 2040
; LENGTH: 931
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MATV03-CLUS34448_1
US-10-767-701-2040

Query Match 24.5%; Score 311.8; DB 18; Length 931;
Best Local Similarity 69.0%; Pred. No. 8.4e-79;
Matches 427; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

Qy 11 GGAGCTCTCAATCAACAACCTGATGATGGCATCTTCTCCGTATGATGATGTTCT 70
Db 313 GCAACTGGTCACTACCTGGCCGACGATGGCATCTTCCATCTTGATGAAGCGGTGCTT 372
Qy 71 TAAAGCAGTGTCTCTCAAGTGCACATCACTTTGGAGGGCAATAGCATGCAAAATTT 130
Db 373 CGAGGCTGTGCTCCCGCGAGCGAGGCAATTTGGAGGGCATCATCATGCCCCACTT 432
Qy 131 GAAGCTCTCTATCACTACCACTGCTGCTGTAGCATACCGGAGGCGGATTTGAAATC 190
Db 433 GAAGCCCCGGTGACCAACAAGGGCGGGCGGTGGAATACAGGGAGGAGATATGAGGCG 492
Qy 191 TTTCCTGTTGATGATTTCAACCTCTTATGACACTTTATTTGACATACCAACAG 250
Db 493 GCTGCGCGCAGAGTACCTTGCAGCCGCTCATGACATGTAACCGACAGCAACACAG 552
Qy 251 TCCATATGAATCAAACTAGAGAGAGAGGAGCCAGTCCATTTGGGGTGAAGTTGATCC 310
Db 553 TCCAGAGAGATCAAGCTCGAAGAAAGATGTATTAATCTTGCTGTGAAGCTATATCC 612
Qy 311 TGCATGTCACAGCAAAATTTCTCAAGATGAGTGAATGATCTTTTCGGGAAGTTTACC 370
Db 613 TGCCTGAGCAATCACTATCCCAAGATGGGTGCTATATATTTGGGAAGTGTGCTGCC 672
Qy 371 AGTTTACAGAAATGTTGAGCATATATGCTCTGCTGGTTCACTGAGAGCTTACTAA 430
Db 673 AGTCCCTCAGAGATGATCAGGCAAGAAATGCAATGCTTTGTCATGAGAGTCAACGA 732
Qy 431 TCCGAGGTTGACATGTTGATAGAGAAAGTATTCATTGAAAGGTTCTTAAGCCGTT 490
Db 733 TCCACATGTTGACACTTGTGACCGGTGAGAGAGTTTTCATATGCAAAATATTTGGCACCACT 792

Qy 491 GGTGCAATATTTCCAAATGGAAGTCTGTATGAGAGATGTTATACCAACCATGATGCTGT 550
Db 793 TGTACAAAACCTCCAGCTGAAAATTGTCATGGAACATATTCACAATATGATGATCAT 852
Qy 551 TAACTTGTGTAATCTTGGACACTGAGGATTTGTTGGACAACTGTGACCCACACATCT 610
Db 853 GAACCTCATGATGATCATGTGAAGAAGTCAATGTTGGCAAGAGTGAATCCAGCATCT 912
Qy 611 TGTTTGAACAGGAATTTCT 629
Db 913 CCTCTCAATAGGAATGCT 931

RESULT 7
US-10-282-122A-41490
; Sequence 41490, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangshu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EPIIIRA.034A
; CURRENT APPLICATION NUMBER: US/10/282, 122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41490
; LENGTH: 1053
; TYPE: DNA
; ORGANISM: Vibrio cholerae
US-10-282-122A-41490

Query Match 22.1%; Score 281; DB 17; Length 1053;
Best Local Similarity 56.8%; Pred. No. 6.9e-70;
Matches 578; Conservative 0; Mismatches 430; Indels 9; Gaps 3;

Qy 15 CTTCATATCAACAACCTGATGATTTGGCATTTATCTCCGTGATGATGATGATGATGATGAT 74
Db 34 CTGACGATATACACCCCAAGACGACTGGACGCTTCACTTGCAGATGGAGATGATGAGCC 93
Qy 75 GCATGTCCTTCAAGAGGACATCATCTTTGGAGGGCAATAGTCATGCCAAATTTGAAG 134

Db 94 GATTACTGTACGGGCATCATCAGCCGTTATTAACGGCGCGCGCTGATCATGCTTAAACCGTT 153

QY 135 CCTCTATCACTACCACTGCTGCTGCTGTAGCATATCCGGGAGGGCGATTTGAAATCTTTA 194

Db 154 CCCCTGTAAACCACTAGATATGCGCTTAACTTATGCTATGCTAGCGCATTTATGGCGCTCAG 213

QY 195 CCTGTGATATGATTTTCAACCCCTTTATGCACTTTATTTGACAGATATACAAACAGTCC 254

Db 214 CC--ACAAGCGCACTTTGAACTTTTGAATGCGCTTAACTCACTCAACGATACATCACTT 270

QY 255 ATGGAATCAAACTAGCAAGAGAGAGCGCAGGTGCTATTTGGGGTGAAGTTTATCCCTGT 314

Db 271 GAGGAATTTGCCAAAGCCAAAGCCTCAGCCAAAGTCGTTACCTAAGCTTTTACCAAGCC 330

QY 315 GGTCCACGCAAAATTTCTCAAGATGAGTGAATGATCTTTTCCGGAAGTGTTAACAGTT 374

Db 331 GGTGACACACCAACTGGATTACGGGTGAC--TTCAAGCAAAAACATTAACCAAGTT 387

QY 375 CTACAAAGAAATGTTGAGCAATTAATATGCTCTGCTGTGTTCAATGGAAGGTTACTAATCT 434

Db 388 TTGCAACGATGACGGAAGTCGGTATGTTGCTTTGTGTGACGGTGAATCTACACGCAAC 447

QY 435 GAGGTGACATGTTTGAATAGAAAAAGTATTCATTTGAACGGTCTTAAGACGTTGAGT 494

Db 448 GAGGTGACATTTTGAACGTTGAAGAAAGACCTTCTTGAACCGTGTTCGCCGATCGT 507

QY 495 CAGAAATTTCCAAATGGAAGTGTGATGATGAGCAATGTTACCAACATTTAGTCTTTAAG 554

Db 508 AATGATTTCCGCACTGMAAAATTTGTCCTTGAACATACACCGCGCATGCACTGAC 567

QY 555 TTGTGTGATCTTGCACTGAAGATTTGTGCAAGCACTGTCACCCCAACATCTTGT 614

Db 568 TTGTGCAA--CAAGCAGCGCATTAACGTTTGGCGCAACGATTAACGGGCACATTTGCTG 624

QY 615 TTGAACAGAAATTTCTCTTCCAAAGGGGCTTAACAACCGAATTAATATGCTTTCAGTC 674

Db 625 TTAAACGTTAAACAATGCTGTTGGCGGATTTGCCCAATTTCTATGTTTCCCAATC 684

QY 675 CTCAAAAGAGATCTCAAGGAGGCACTTGTGCACTGTATCAAGTGAAGTGAAGA 734

Db 685 CTAAAGCGTCACTACCAACACCGTTATGTGCGAGCGGCACTCTGGAGCAAGAAA 744

QY 735 TTTTTCCTTGGGACTATATGCTCTCATGATTTGAAGAAAGAAAGTGTCTTGTGGA 794

Db 745 TTCTTCTTAAGTAACGACTGTGCCCGGACGCAAAAGCCGCAAAAGGCCGCTTGGGC 804

QY 795 TGGTGAGTATTTAAATGCACTGTATGCTTGTCAATATATGCAAGAGTGTGGAAG 854

Db 805 TGGCAGGTTCTTAACAGCCCAAGCAGCTTGAATGTATGCGGAAGTGTGAAAA 864

QY 865 GAAAAATGCACTGCAAGAGCTTGAAGCAATCACTGATGCTCAATGGAACCAATTTTATGG 914

Db 865 GAAAGTAAAGCTGAAGAAATCTGGAAGCGTTGCTATGCTTTATGACCTGATTTTCTACGCG 924

QY 915 CTCTCTAGGAACAATCTAAAGATTAAGTTAAGTAAGACGCATGGAAGTAAACCGCAATCC 974

Db 925 CTACTCTGCAATCAAGAGACCGTCAACGCTCAACCAAGAACCTGGCAGTCTGCAAGAAAC 984

QY 975 TTTTCTTATGCACTAGAGATATTAATTTCCATGTTTGTGCTGGAATATGCTGCACTGG 1031

Db 985 ATGCGCTTTGACAGCGATATCGTGTGCTCAATCCGCGCGGGTGAATAATTAATCAAGTGG 1041

```

1  APPLICANT: Zykind, Judith
2  APPLICANT: Wall, Daniel
3  APPLICANT: Trawick, John
4  APPLICANT: Carr, Grant
5  APPLICANT: Yamamoto, Robert
6  APPLICANT: Forsyth, R.
7  APPLICANT: Xu, H.
8  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
9  FILE REFERENCE: RITTA.034A
10 CURRENT APPLICATION NUMBER: US/10/282.122A
11 CURRENT FILING DATE: 2003-02-20
12 PRIOR APPLICATION NUMBER: 60/191,078
13 PRIOR FILING DATE: 2000-03-21
14 PRIOR APPLICATION NUMBER: 60/206,848
15 PRIOR FILING DATE: 2000-05-23
16 PRIOR APPLICATION NUMBER: 60/207,727
17 PRIOR FILING DATE: 2000-05-26
18 PRIOR APPLICATION NUMBER: 60/230,335
19 PRIOR FILING DATE: 2000-09-06
20 PRIOR APPLICATION NUMBER: 60/230,347
21 PRIOR FILING DATE: 2000-09-09
22 PRIOR APPLICATION NUMBER: 60/242,578
23 PRIOR FILING DATE: 2000-10-23
24 PRIOR APPLICATION NUMBER: 60/253,625
25 PRIOR FILING DATE: 2000-11-27
26 PRIOR APPLICATION NUMBER: 60/257,931
27 PRIOR FILING DATE: 2000-12-22
28 PRIOR APPLICATION NUMBER: 60/267,636
29 PRIOR FILING DATE: 2001-02-09
30 PRIOR APPLICATION NUMBER: 60/269,308
31 PRIOR FILING DATE: 2001-02-16
32 Remaining Prior Application data removed - See File Wrapper or PALM.
33 NUMBER OF SEQ ID NOS: 78614
34 SOFTWARE: PatentIn version 3.1
35 SEQ ID NO 42220
36 LENGTH: 1047
37 TYPE: DNA
38 ORGANISM: Yersinia pestis
39 US-10-282-122A-42220

```

RESULT 8
US-10-282-122A-42220
Sequence 42220, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liansu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlisen, Karl

	Db	Qy	Db	Qy	Db	Qy	Db	Qy
262	AAAGAGTTGACCA CGGGTTTGTGACCAAGCGCTTTTACCGGGCCAACTGTATCCGCC	3262	315	GGTGCCACGACAATTTCTCAAGATGAGTGA CTGATCTTTTCGGGAAGTGTTCACGATT	3745	322	AATGCCACCACTCACTCACTGAGGTGATCTGCACAT---CCCGGCAATTTACCGGTTG	3788
315	GGTGCCACGACAATTTCTCAAGATGAGTGA CTGATCTTTTCGGGAAGTGTTCACGATT	3745	322	AATGCCACCACTCACTCACTGAGGTGATCTGCACAT---CCCGGCAATTTACCGGTTG	3788	375	CTACAAGAAATGCTTGAGCATTAATATGCTCTGCTGGTTCATGAGAGCTTACTAATCCT	4344
322	AATGCCACCACTCACTCACTGAGGTGATCTGCACAT---CCCGGCAATTTACCGGTTG	3788	375	CTACAAGAAATGCTTGAGCATTAATATGCTCTGCTGGTTCATGAGAGCTTACTAATCCT	4344	379	TTTGACCAAAATGCAAAAGATAGGCATGCCCTCTTATTACCGGTGAGTAAACAAGTCCG	4388
375	CTACAAGAAATGCTTGAGCATTAATATGCTCTGCTGGTTCATGAGAGCTTACTAATCCT	4344	379	TTTGACCAAAATGCAAAAGATAGGCATGCCCTCTTATTACCGGTGAGTAAACAAGTCCG	4388	435	GAGGTGACATGTTTGAATAGAGAAAAGTATTATTGAAACGGTTCTAAGACCGTTGGTG	4944

Db 439 GCGGTTGACATCTTTTATGTGTGAAGCCGCTTTTATTGACCAATTTTAGAGCCCATTCGC 498

Qy 495 CAGAAATTTCCACAAATTGAAGGTCGTGATGAGCATGTTACCAACANTGATGCTGTTAAG 554

Db 499 CAAAGTTTCCGACCTAATAAATTGTCTTTGAGCATATCAAGACCAAGAGTCGGCAGAT 558

Qy 555 TTTGTGAAATCTTGCACTGAAGAGATTGTTGCAGCACTGTACACCCCAACAACCTTTGTT 614

Db 559 TATGTGCTGACGACA--ATCGTTTCCTTGGGGCAACCGTCAACGCCCAACACTTATG 615

Qy 615 TTGAACAGAAATTCCTCTTCCAAAGGGGCTTACACCGCATTAATTATGCTTCCTGATC 674

Db 616 TTTAACCGCATCACATGCTGTGAGCGGTATTCGCCCACTTGTTCTGCTCCATA 675

Qy 675 CTAAAGAAGATTCACAGGGAGGCACTGTGTGCTAGCTGTAAACAAGTAAAGAA 734

Db 676 TTGAAGCGACACCCATCAGCAAGATTCGCCGACCGCTGCCAGTGTCTGTATCGC 735

Qy 735 TTTTTCCTTGAGCATGTAGTGTCTCTCATGATAGCAAGAAAGATGTTCTTGTGGA 794

Db 736 TTCTTCTTGGGACCGATTCAGTCCCAATGCCAAACATCGTAAGATATCTTTGGGCG 795

Qy 795 TGTGCTGTATTTACATGACACCTGTAGCCTTGTCAATATGCAAGAGTGTGAAAG 854

Db 796 TGTGCGGGTATTTCAACGCCCGACGGGATTCGCTTATGCTTCGTTGTAAGGA 855

Qy 855 GAAATGCACTCGACAAAGCTTGAAGCATTCAGTCTCAATGACACGAATTTTATGGG 914

Db 856 CTGAATCATTTGCAACATCTGGAAGGCTTTTGGCCTTAAATGGGCCACGAATTTATGGC 915

Qy 915 CTTCCTAGAACAACTCAAGATTTAAGTTGAGTAAGACGCATGGAAGGTACCGGAATCC 974

Db 916 TTGCTGTATATGATGACGTGTGTGAATGGTTGCACTCAATCTCTGACGACGAAGAG 975

Qy 975 TTTTCTTAATCAGAGATATTTATTCATGTTTCCATGTTGCGGAATGCTGCATGTG 1032

Db 976 ATCCCATTTAGCAATGATGCGTATATTCCTTTTCCTTGGGGTCAAAAGCTTAATTTGGT 1033

```

RESULT 9
US-10-282-122A-8516
/ Sequence 8516, Application US/10782122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haeselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zysekind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsythe, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282.122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/

```

	Prior Application Number:	60/253,625
	Prior Filing Date:	2000-11-27
	Prior Application Number:	60/257,931
	Prior Filing Date:	2000-12-22
	Prior Application Number:	60/267,636
	Prior Filing Date:	2001-02-09
	Prior Application Number:	60/269,308
	Prior Filing Date:	2001-02-16
	Remaining Prior Application data removed - See File Wrapper or PALM.	
	Number Of SEQ ID NOS:	78614
	Software:	Patentin version 3.1
	SEQ ID NO:	8516
	Length:	1032
	Type:	DNA
	Organism:	Acinetobacter baumannii
	US-10-282-122A-8516	
Query Match	21.0%	Score 267.4; DB 17; Length 1032;
Best Local Similarity	56.4%;	Pred. No. 5.7e-66;
Matches	540; Conservative	0; Mismatches 411; Indels 6; Gaps 2
OY	27	CACCTGATGATTGGCATCTTCATCCGTATGSGTATGTTCTTAAGGCAAGTGCTCT 86
DB	22	CAGCCAGATGATTGGCAGTGCACATTTAGGTATGTTGAATTAACGACTGTCCA 81
OY	87	CACAGTCACATCACTTTGGGAGGSCATPAGTCAGCCAAATTTGAMCGCTCTATCACT 146
DB	82	GATTTGGCTTAAACAATTGCCCGTCGATTTGTATGCTTACCCTGTAOCGCTGTAAA 141
OY	147	AACCATGCTGCTGCTGTAAGCATACCGGAGGCGATATTGAATCTTTAAGCTGTATAGT 206
DB	142	ACAGTAGAAGAAGCTTTAGCTTATGCGAACGCACTTCTGCTCATGTTCCGGAGGCAT 201
OY	207	GATTTC AACCTCTTATGACATTTATTTGACAGATTAACAACAGTCTATGGAAATCAA 266
DB	202	AAATTTGACCTCGATGATGTTATTTATTTATGACCACTTCAACAGATBAAGTTGCT 261
OY	267	CTAGCAAGAGAGAGCCAGTGGTATTTGGGGTGAAGTTGTAACCTGCTGTCAGCACGA 326
DB	262	AAATTAAGAATCGAAGCATGTATAATGCGATTAAAGCTTATCTGCTGGTGCACACA 321
OY	327	AAATTCGAAGTGAAGTGAAGTATCTTTTCGGGAAGTTTACCAAGTTCTACAGAATG 386
DB	322	AACTCGATTAATGGTGTAGTATATTT--CGTAAAGTTTATGACAGTCATTAAGCATTA 378
OY	387	GTTGAGCATATATATGCTCTCTGCTGTTCTATGAGAGAGTTTACTTAATCCGAGGTGCATG 446
DB	379	GAAAGGACATCAAGTCCCGTATTTGCTTCAATGGTGAAGTAACTACACATATCATGATGATTT 438
OY	447	TTGTATGAGAAAGATATTCATTGTAAGACGGTTCTTAAGACCGTGGTGAGAAATTTCCA 506
DB	439	TTTGATCGGAAAAACGCTTCTCAATGAGGATATTTATCACTCTATTTGAACAGTCCCG 498
OY	507	CAATTGAAGTGTGATGAGAGCATGTATACACATTTGATGCTTTAAGTTGTTGATCT 566
DB	499	AAACTTAAGTTGTGCTTGAGCACATCACTACAGATGAGACACACTTGT--TTA 555
OY	567	TGCACTGAAGATTTGTTGCACAACTGTCACCCACACATCTGTTTTGAACAGGAT 626
DB	556	GAAACAAGACCGTAATGTGGCGCAACAATCACTCCACACATTTATTTAACCGTAAT 615
OY	627	TCTCTCTTCCAGAGGGGCTTACAACCGCATTAATTAATCTGCTTCCAGTCTCTCAAMAAGAG 666
DB	616	GATATGTTGGTGGTGATTTAAACCGCACTTTACTGTTTACAAATTTTAAAGCTCAA 675
OY	687	ATCCAAGAGGAGCATTTGTGTCAGCTGTAAACAGTGAAGTAAAGATTTTCTTGGG 746
DB	676	ACACATCAAAACGACTTTCCTTGAAGTGGCGAAGGCGGTAACTCTAAATTTTCTTAGGT 735
OY	747	ACTGATAGTCTCCATCATATAGAGCAAGAAAAGAGTCTTCTGAGATGCTGCTAT 806
DB	736	ACAGAGAGGCTCTCATACACAAATGCAAAAGAGATGCTTGTGATGTGCAAGCTCG 795

QY 807 TACATGACCTGTAGCTTGTGATATGCGAAGGTGTTGAAAAGAAATGCACTC 866
DB 796 TATAGTGACCAATATGATTAAGCTTTATGACAAAGATTTGACNAGTGGTAAATTA 855
QY 867 GAAAGCTTGAAGCAATTCATGACTTCAATGAGCAGATTTTATGAGCTTCCAGAAC 926
DB 856 GAGCGTCTGAAAGTTTGTCTAGCCATTTTGGTGGAGCTTACGGCTTACAGCTAAT 915
QY 927 AACTCAAGATTAACTGATGAGAGCCATGGAAGGTACCGGATCTTTCTTAT 983
DB 916 ACTTTCACCATTAATCTTGGTTAAAGAAATACCTCGTTCCAGAACTTTTGAATAT 972

RESULT 10
US-10-282-122A-14242
; Sequence 14242, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14242
; LENGTH: 1056
; TYPE: DNA
; ORGANISM: Burkholderia mallei
US-10-282-122A-14242

Query Match 20.9%; Score 265.2; DB 17; Length 1056;
Best Local Similarity 54.4%; Pred. No. 2.5e-65;
Matches 559; Conservative 0; Mismatches 463; Indels 6; Gaps 1;

QY 14 GCTCTCAATCAACAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 73
DB 24 GCTGACCTCTGCG 83
QY 74 GCGAGTGTCTCTCAAGTGCATCACTTTGGAGGCGCAATAGTCAATGCCAATTTGAA 133

DB 84 GCGCTCTCTGCGCGCACCGCGCTCGCAGATTGCGCGCGCGCGCATCATGCGCAACTGAA 143
QY 134 GCTCTCTATCACTACCACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 193
DB 144 GCGCGCGCTGACGACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 203
QY 194 AACTGTGATGTATTTCAACCTCTTATGACACTTTTATGACAGATTAACAACAGATTC 253
DB 204 GCG 263
QY 254 TATGAAATCAAACTAGACAG 313
DB 264 CGAGCAATTCG 323
QY 314 TGGGCGACAGCAAAATTTCTCAAGATGAGTGAATGATCTTTTGGAGAGTGTATACAGT 373
DB 324 GGGGCGACGACCAACCTCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 383
QY 374 TCTACAAGAAATGTTGAGCATATATATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 433
DB 384 GCTCGAGGGGATGACAGAAAGTGGGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 443
QY 434 TGAAGTTGACATGTTGATGAGAAAGATATTCATGAAACGCTTCTTAAGACCGTTGT 493
DB 444 GTGATGACCTGTTGACCGCGAGAAAGTGTTCATGATGCGCTGATGAGCGCGTGG 503
QY 494 GCAGAAATTTCCAAATTTGAAGGTCGTGATGAGACATGTTACCAACATGATGCTGTTAA 553
DB 504 CCGCGCGCTGCG 563
QY 554 GTTTGT-----TGAAATTTGACTGAGAGATTTGTTGCGACCACTGTCCACCCCAACA 607
DB 564 CTACGTGCGCATGCGACG 623
QY 608 TCTTGTGTTGAACAGAAATCTCTCTTCCAGGGCGCTTACCAACCGCATTAATTAAGTCT 667
DB 624 TCTGCTGTAACACCGCAATGCAATGTTTTCGGCGGATCTCGCCACATTAATTAAGTCT 683
QY 668 TCCAGTCTCTCAAAAGAGATTCACAGGAGGCACTTGTGTAGCTGTAAACAAGTGAAG 727
DB 684 GCGGTCTCAAGGCGAGAGCGATCGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 743
QY 728 TAAAGATTTTCTTGGGACTGATGATGCTCTCTCTATGATTAACAAGAAAGAGTGTTC 787
DB 744 TCGCGCTTCTCTCTGCGACCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 803
QY 788 TTGTGATGTGCTGTATTTAACAATGACCTGTAGCTTGTGATATGAGAAAGTGTTC 847
DB 804 GTGGGCTGCGCGCGCTGCTACACCGCGCTGCGCGCTGCGCGCTGCGCGCGCGCGCG 863
QY 848 TGAAGAGAAATGCACTGACAGAGCTTGAAGCAATTCATGCTTCAATGAGCAGATTT 907
DB 864 CGACAGAGGCGGCGCGCTGACAACTGAAAGCTTTCGCGAGCTTCTTTCGCGCGCGCG 923
QY 908 TATAGGCTTCTTGAACAACCTCAAGATTAACTGATGAGAGAGAGAGAGAGAGAGAG 967
DB 924 CTACGCTTTCG 983
QY 968 CGAATCTTTTCTTATGATGATGAGAGATTAATTTCCATGTTTGTGTGTAATGCTGCA 1027
DB 984 GCGCGAGATCGACG 1043
QY 1028 CTGCTGTC 1035
DB 1044 CTGCGCGC 1051

RESULT 11
US-10-282-122A-11985
; Sequence 11985, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haseibeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EUTRA.034A
CURRENT APPLICATION NUMBER: US/10/282.122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11985
LENGTH: 1092
TYPE: DNA
ORGANISM: Burkholderia cepacia
US-10-282-122A-11985

Query Match 20.8%; Score 264.4; DB 17; Length 1092;
Best Local Similarity 56.2%; Pred. No. 4.3e-65;
Matches 540; Conservative 0; Mismatches 411; Indels 9; Gaps 2;

14 GCTCTCAATCACACAACTGATGATGGCATCTTCATCTCGTGATGTGTTCTTAA 73
63 GCTTTCGCTCGCCGTCGCGACGACTGCACTGCGAGACGCGCATGTCTGC 122
74 GGCAGTTGTCTCTACATGACATCACTTTGGAGGGGCAATGATGCAAAATTTGAA 133
123 GCGCGTGTGCGGACACGCGCCGCAATGCGCGCCCAATCGCATGCGCACTGAA 182
134 GCTCTCTATCACTACCACTGCTGTGTAGACATACCGGAGGGGATATTGAAATCTTT 193
183 GCGCGCGTCAAGACACCGCGGAGCGCGCATGATGCGGATCTCTGCGCGCT 242
194 ACCTGTGATAGTGAATTTCAACCTCTTATGACATTTATTTGACAGATCAACCAATCC 253
243 GCGCGCGGAGTGAAGTTGAGCCGCTATGACCTGTATACCTGACGACCAACCGCCGCC 302
254 TATGAAATCAAACTGAG 313
303 CGACAAATTCGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
314 TGTGTCCACGACAAATTTCTCAAGATGAGTGAATGATCTTTTTCGGGAAAGTGTTCAGAT 373
363 AGGCGCCACGAGAGATTCGACCAATGCGGTCAACGATC---TCGGGAATGTGCGAAGAC 419
374 TCTACAGAAATGTTGAGCATATATGCTCTGCTGTTCTATGAGAGGTTACTAATCC 433

420 GCTCGAGGCGATGACAGAAACCGGCATGCCGCTGTCTGTCACGCGAGAGTACCCGATGC 479
434 TGAAGTTGACATGTTGATATGAGAAAGGATTCATTTGAAGGCTCTAAGACGTTGGT 493
480 GTTCATGACCTGTTGACCGCGAGAGAGGTTTCATGACCGCGTATGACCGCTGCG 539
494 GCAGAAATTTCCAAATTTGAGAGTGTGATGAGCATGTTACCAATGATGATCTGTAA 553
540 CCGCATTTCCGAGGCTTGAAGTCTGTTCGAAACATACATACGAGAGAGAGAGAGAGAG 599
554 GTTTGT-----TGAATCTTGCACCTGAAGGATTTGTCAGAACCTGTCAACCCACACA 607
600 CTACGTCGCGAGAGCGGAGCGCGCGCGCGCTGTCTGCGGCGAGATCAACCGCGATCA 659
608 TCTTGTGTTGACAGAAATTTCTTCTTCAAGGGGCTTCAACCCGATTAATACGCT 667
660 CTGCTGTACAAACCGAACCGCGCTGTCTGTCTGCGGAGATCCGCCGATTAATCACTGCT 719
668 TCCAGTCCCTCAAAAGAGATTCACAGGAGGACCTGTGTCACTGTATCAAGTGAAG 727
720 GCGGTGCTGAGCGGAGAGCGATGCTGTGCGCTGTGTGAGAGCCGCGACCTCGGCA 779
728 TAAAGATTTTCTTGTGGACTGATAGTCTCTCATGATGACGAAAGAGTGTTC 787
780 CCGCGCTTCTTCTCGGACCGACAGCGCGCGCGACGCGCGAGAGAGAACCGC 839
788 TTGTGATGTCTGTGATTTTACATGACCTGTAGCTTGTCAATATATCGAAGTGT 847
840 GTGCGGTGCGCGGCTGTACACGCGCTGTGACGCGCTGTGATGATGACGAGAGGCT 899
848 TGAAGAGAAATGACATCGACAACTTGAAGCTTCACTAGCTTCAATGAGACGAGATTT 907
900 CGACAGCGCGCGCGCTGTACAGAGCTGAAAGGTTGCCAGCTTCTTCCGCGCGATTT 959
908 TTAATGAGCTTCTTGAAGAACATCAAGATTAATGATGATGAGAGCGCATGAGAGTACC 967
960 CTACGCGCTCGCGGACGCGCGCGAGAGATGATGATGATGATGATGATGATGATGATGAT 1019

RESULT 12
US-09-790-988-1
Sequence 1, Application US/09790988
Patent No. US20020127687A1

GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: WATANABE, HIDEKI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 20.8%; Score 264.2; DB 9; Length 640681;
Best Local Similarity 54.0%; Pred. No. 2.2e-63;
Matches 564; Conservative 0; Mismatches 478; Indels 3; Gaps 1;

20 AATCACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 79
368078 AATCATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 368137
80 TGTCTCTACAGTGCATCACTTTGGAGGCAATAGTCAATGCCAAATTTGAAGCTTCC 139

Db 385 CGCATGAAAAAATCCGTATGCGCTACTGTGATGTAAGTGAACATGCAATGATATC 444
Qy 441 GACATGTTGATAGAGAAAAGTATTTCATTGAAACGGTTCTTAAGCCGTGTGCAGAAA 500
Db 445 GACATTTTGTATGTGAAGCGCCTTTATGAAAAGGTGATGAACCTCTGCCGACGGC 504
Qy 501 TTTCACAAATTGAAGTCTGTATGAGACATGTTAACCACTTGTATGCTTTAAATTGTT 560
Db 505 CTGACTGCGCTGAAAGCTGTTTGTGAGCATCACCAACCAAAATGTGCGCATATGTC 564
Qy 561 GAATCTTGCACTGAAGATTTGTTGACGAACGTGCCCCACAACTCTGTTTGAAC 620
Db 565 CTGACGGAATATAGC--GCTGGCTGCCACATCACTCCGACGATCTGATGTTTAA 621
Qy 621 AGGAATCTCTCTTCCAGAGGGGCTTACAACCGCATTAATTAAGCTTCCAGTCTTCAA 680
Db 622 CGCAACCATATGCTGTGTAAGCGCTGCTCCGACCTGATTTGTCTACCATCCTCAAA 681
Qy 681 AGAAGATCCACAGGAGGACCTTGTGTCACTGTAAACAAGTGAAGTAAAGATTTTTC 740
Db 682 CGTAATATTCACCAACAGGACATTCGTAATGCTGTCGCCACGGTTTAAATCGATATTC 741
Qy 741 CTGGGACATGATGTGCTCTCATGATAGAGAAAGAAAGATTTCTTGTGATGTC 800
Db 742 CTGGATACGATTTCTGGCCACATGACGTCATCGCAAGAGACAGTTGCGGCTGCGC 801
Qy 801 GGATTTTCAATGACACCTGTAGCCTTGTCAATATATCGAAGGTTTGAAGAAAT 860
Db 802 GCGCTCTTCAACGCCCAACCGCGCTGGGACGTTACCTTCTTGAAGAAATGAT 861
Qy 861 GCACTGACAGAGCTTGAAGACATTCATAGCTTCAATGACACAGATTTTAAATGCGCTTCT 920
Db 862 GCTTTGACAGCTTTGAAGCATTTCTGTGTGTAACGGCCGACGTTCTATGAGGTTCCGC 921
Qy 921 AGGAACAATCTCAAGATTTAACTTGAATGAAGCCGCAATGAAAGTAACTCTTTTCT 980
Db 922 GTCAACACACATTCATCGAATCTGTACGTGAAGAGACAAACGTTGTGAAAGATGCA 981
Qy 981 TATGATCAGAGATATTAATCCCATGTTTGTGCTGTAATGCTCGACTGCT 1032
Db 982 CTGACTGATGACACGCTGTGCTCAATCTCTGCGCGGGAACGTAACCTGCT 1033

RESULT 14
US-10-282-122A-30305
Sequence 30305, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EUTRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patentin version 3.1
SEQ ID NO 30305
LENGTH: 1047
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30305

Query Match 19.0%; Score 241; DB 17; Length 1047;
Best Local Similarity 53.7%; Pred. No. 2,4e-58;
Matches 546; Conservative 0; Mismatches 465; Indels 6; Gaps 2;

Qy 15 CTCTCATGACACAACTGATGATGTCATCTTATCCGTATGATGATGTTCTTAAG 74
Db 13 CTACCCCTCTCTGCGCCCGACGCTGGCAATTCATCTGATGCGGTGCGCGCTGCC 72
Qy 75 GCAATGCTCTCTCAGATGACATCACTTTGGAGGGCAATGATGCAAAATTTGAG 134
Db 73 AATACGTCGGGACACGCGCCGACCTTGGCGCGCCATGTCATGCGGAACCTGCTG 132
Qy 135 CTTCTATACATCACTCACTGCTGCTGTGATGATCCGGAAGGCAATTTGAAATCTTTA 194
Db 133 CCGCGGTGGGCAACGCGCCGAAAGCCGATACCGGAGGCGCATCTCGCCCGCT 192
Qy 195 CTTGTGATGATGATTTCAACCTCTTATGACATTTATTTAGACAGTATACACAGTCT 254
Db 193 CCGCGCGGCGCGCTTGAAGCGCGCTGATGATGATGATGATGATGATGATGATGATG 252
Qy 255 ATGGAATCAAACTAGCAAGAGAGCAGATGCTATTTGGGAGTGAATTTGATCTCTCT 314
Db 253 GAGGAATCCGTACCGCCGCAAGCGCGCTGATGATGATGATGATGATGATGATGATG 312
Qy 315 GGTGCCACGACAAATTTCTCAAGATGATGATGATTTTTCGGAAGTCTTACAGTT 374
Db 313 GGGCGCACACCAACTCGACTGCGGTGAC--CCGCAATCGACAAATATCTTGAAAGCG 369
Qy 375 CTACAAAGATGTTGATGATATATGCTCTGCTGTTGATGAGAGGTTACTAATCTT 434
Db 370 CTGAGGCGATGCGCGAGTCTGCGATCCGCTGATGATGATGATGATGATGATGATG 429
Qy 435 GAGGTGACATGTTGATGAGAAAGTATCTTGAACGTTCTTAAGACCGTTGGTG 494
Db 430 GAGGTGACATGTTGATGAGAAAGTATCTTGAACGTTCTTAAGACCGTTGGTG 489
Qy 495 CAGAAATTTCCAAATGAAAGTGTGATGAGAGATTTTACACATTTGATGCTGTTAAG 554
Db 490 GAGGCTTCCGACCTCGAAGGTGCTTGAAGACATCACACCGGGAAGCGCCGACG 549
Qy 555 TTTGTTGAATCTTCACTGAAGATTTGTTGACGAACCTGACACCCACCAATCTTGT 614
Db 550 TTCTGTCGGGAAGCCC--GGCCAAAGTGGCGGACCAATTTACCGCCCATCTGCTG 606
Qy 615 TTGAACGGAATTTCTCTTCCAGGGGCTTACAACGCGATTAATTAATCTGCTTCAAGT 674
Db 607 TACAACGCAACCACTGCTGCTGGGATTCCTGCTGACCTTATGCTGCTGCTGCTG 666
Qy 675 CTCAAAAGAGATCCACAGGAGGACCTTGTGCTGATGATGATGATGATGATGATG 734
Db 667 CTCAAGGCAACCAACAGGAAGCCCTGTGACGCGCGGTGAGCGGCAATCCGAA 726

QY 735 TTTTCTTGGAAGCTGATAGTCTCTCTATGATGACGAAGAAAGTGTCTTGTGA 794
DB 727 TTCTTCTCGACAGCTCGCGCCGACCGCCGCAAGCAAGCCGCTGGGT 786
QY 795 TGTGTGATTTTACATGACCTGTAGCTTGTGAGTATATGGAAGTGTGAAG 854
DB 787 TGGCGGGCTGTCTACAGGGCTTACGGCCATCGAGCTGTATGCGAGGCTTGAACG 846
QY 855 GAAATGACCTCGAAGCTTGAAGCATTCATGACTTCAATGAGACCAATTTTATGG 914
DB 847 CGCATGCGCTGGAAGAGTGAAGGCTTGCACGCTTGCACGCGCCGACTTCTATGCG 906
QY 915 CTCTTATGGAACAATCAAGATTAAGTGAAGAGCGCCATGGAAGTATCCGATCC 974
DB 907 CTGCGGCGCAACCGACCGGATTCACCTGTCTCGGAGAAATGCGAGCGCCGCGAC 966
QY 975 TTTTCTATGATCAGAGATATTTATCCATGTTTGTGTGGAATGCTGCATCG 1031
DB 967 CTCCCTTCGCGCATTCGACGTGTGCGCTGCGCGCGAGAGACACTCCGCTGG 1023

RESULT 15

US-10-282-122A-15202
Sequence 15202, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Heselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Wall, Daniel
APPLICANT: Zyskind, Judith
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15202
LENGTH: 1062
TYPE: DNA
ORGANISM: Bordetella pertussis
US-10-282-122A-15202

Query Match 18.7%; Score 238.2; DB 17; Length 1062;
Best Local Similarity 53.4%; Pred. No. 1.5e-57;

Matches 554; Conservative 0; Mismatches 468; Indels 15; Gaps 2;
QY 12 GAGCTCTAATCAACAACTGATGATGAGCATCTTCAATCTCCGTATAGTATGTTCTT 71
DB 22 GAATTCACATACCGCCCGGACGATGAGCACTTCACTGGGCAATGAGCGCCCTC 81
QY 72 AAGGACATGCTCTCACTAGTCATCATCTTTGGAGGCAATAGTCGCAAAATTTG 131
DB 82 GAGCGGTGTGTGATGATACGCGCGGAGGTCAGTTCGCGCGCCATGTGATGCTCAACTC 141
QY 132 AAGCTCTTATCACTACCACTGCTGCTGTGATAGCATACCGGAG-----GGCG 179
DB 142 CGCCCGCGGTGACACACACAGAGGCGCTGTATACGCGCGCGGATGACAGCGCGG 201
QY 180 ATATGAATCTTTACTCTGTATGATGATTTTCAACCTCTTATGACATTTATTTGACA 239
DB 202 CTGCGGAGCGGCGGATGACGCTGCTCTTCAAGCTTGAATGACGCTGATACCTGACC 261
QY 240 GATACAAACAGTCTTATGGAATCAAACTAGCAAGAGAGCCAGGTGTTTGGGGTG 299
DB 262 GACCAACCGCGCGGAGGAAATCGTGCAGGATGATGCTCGGAGAGTGCAGCGGTC 321
QY 300 AAGTTTACCTGCTGTGTGTCACGACAAATTCGAAATGAGTGAATGATCTTTTGGG 359
DB 322 AAGCTGTATCCGCGCGGCGCACCAAACTCCGATGCGCGGTGACCGACTGTCTGGC 381
QY 360 AAGTGTTCACAGTTCTCAAGAAATGTTGAGCATATATGCTTGTCTGTATGATGA 419
DB 382 AAGTGGCGCGCGCGCTGCTGCTGCTGATGAGCGCTGCGGATGCTGCTGTGTGACGGC 441
QY 420 GAGGTATCAATCTGAGAGTTGACATGTTGATGAGAAAGATATCATTTGAAACGTT 479
DB 442 GAGGTACCGATCCGCGCATGACATGATGACCGGCAAGCGGTTCATGACAGTGC 501
QY 480 CTAAAGCGTGTGTCAGAAATTTCCAAATTAAGTGTGATGAGCATGTTTACCC 539
DB 502 ATGACGCGCTGCGCGCGCTGCTGCTGCTGATGAGCGCTGCGGATGCTGCTGTGACGC 561
QY 540 ATGATGCTGTTAAGTTGTTGAATCTTGACATGAGATTTGTCAGCAATGTCACCC 599
DB 552 CGGAAAGCGCCCACTAGT---GCGGACGCGGAGAGAGCCACCGCCGACATCAAG 618
QY 600 CCACAAATCTGTTTGAACAGAAATTCCTCTTCCAAAGGGGCTTACACCGCATAT 659
DB 619 CCGAGACATGCTCTCAACCGCATGCAATTTTACGCGGCGCGCTGCGCGCATTTGG 678
QY 660 TACTGCTTCAAGTCTCAAAAGAGATCCACAGGAGCACTTGTGACGTGATACA 719
DB 679 TATTGCTTGCAGATCTCAAGCGGAGTGATCGCAGGCGCTGTGAGGCGGCAC 738
QY 720 AGTGAAGTAAAGATTTTCTTGGAGTATGATGCTCTCATGATGAGAGAA 779
DB 739 AGCGGAGTCCGCTTCTTCTGAGTACGACAGCGGCCCAAGCGCGCGCTGCAAG 798
QY 780 GAGTGTCTTGTGATGCTGTGATTTTCAATGACCTGTAGCTTGTGATGATGAG 839
DB 799 GAACATGCTGTGTGCGCGGCTGTATAGGCTGTGATGATGATGATGATGATGATG 858
QY 840 AAGGTGTTGAAGAAAGAAATGACATCGACAGCTTGAAGATTCATGATGATGATG 899
DB 859 ACGCGCTGACGCGGTGCGCGCTGATCGGTGAGAGCTTTTGCAGCTTCAAGCG 918
QY 900 CCAATTTTATGAGCTTCTAGAACTCAAGATTAAGTTGAAGTGAAGAACCCATGG 959
DB 919 CCGGATTTTACGGCTGCGCGGCAATAGGCTACGCTGCGCGGAGAGCTTAC 978
QY 960 AAGGTACCGGATCTTTTCTTATGATCAGAGATATATTCATGTTGTGCTGATGA 1019
DB 979 GAAATTCGCGGAGTGTGCTGTGCGGATGACACCTGTGCGCTGTGCGGCGGAG 1038
QY 1020 ATGCTGACGTGTTGCC 1036
DB 1039 TCGCTCGCTGCGCGGCG 1055

Mon Feb 14 10:10:56 2005

us-10-070-277-1.rnpb

Page 13

Search completed: February 12, 2005, 07:01:22
Job time : 817 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 12, 2005, 01:51:56 ; Search time 4610 Seconds
(without alignments)
10494.509 Million cell updates/sec

Title: US-10-070-277-1

Perfect score: 1271

Sequence: 1 tgcgaataatgagctctca.....tgaagattcagatataaaa 1271

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hlc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gsa1.*
9: gb_gsa2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	816.8	64.3	820	5	B0516583
2	767.4	60.4	1007	7	CK252266
3	744.6	58.6	751	5	B0516584
4	690.8	54.4	915	7	CK253198
5	689.6	54.3	790	7	CK242513
6	662.6	52.1	920	7	CK250612
7	658.2	51.8	825	7	CK249779
8	616.2	48.5	723	7	CK242510
9	607.6	47.8	708	7	CK251510
10	599.4	47.2	732	4	B1927946
11	595	46.8	957	7	CK242511
12	585.4	46.1	1301	3	CNS0A31G
13	583.8	45.9	1294	3	CNS0A38S
14	574.4	45.2	1297	3	CNS0A2XU
15	561	44.1	792	7	CK251764
16	555.2	43.7	1300	3	CNS0A3CO
17	550.8	43.3	1080	9	CL959705
18	549.6	43.2	583	1	A1895210
19	537.4	42.3	564	2	BE451410
20	514.4	40.5	532	2	BE451489
21	506.2	39.8	519	7	CK253199
22	502.8	39.6	871	7	CK247301
23	500	39.3	954	7	CK248473
24	498	39.2	1296	3	CNS0A4FR

25	496	39.0	905	7	CK246423	CK246423	EST730060
26	481	37.8	821	7	CK242512	CK242512	EST726149
27	465	36.6	957	7	CK245528	CK245528	EST729165
28	462.8	36.4	897	7	CK249778	CK249778	EST733415
29	441.4	34.7	1145	7	CK172140	CK172140	FGAS04769
30	441.2	34.7	641	4	BI203471	BI203471	EST521511
31	438.2	34.5	1153	7	CK170617	CK170617	FGAS04556
32	423	33.3	834	7	CK245538	CK245538	EST729175
33	421.4	33.2	1137	7	CK171201	CK171201	FGAS04635
34	404.2	31.8	743	5	BU636341	BU636341	049G06 1n
35	403.2	31.7	887	7	CV130150	CV130150	MdFrc088
36	398.8	31.4	410	7	CK245539	CK245539	EST729176
37	370.2	29.1	1179	7	CK171534	CK171534	FGAS04683
38	365.6	28.8	1156	7	CK169818	CK169818	FGAS04446
39	362.2	28.5	1056	7	AY109986	AY109986	zee. may8
40	360.8	28.4	1174	7	CK171933	CK171933	FGAS04741
41	353	27.8	627	7	CF445164	CF445164	EST681509
42	346.8	27.3	607	2	BE344806	BE344806	946029B05
43	344.6	27.1	758	7	CV084527	CV084527	MdFrc3085
44	336.8	26.5	1093	7	CK171050	CK171050	FGAS04615
45	329	25.9	786	4	BU576697	BU576697	BU576697

ALIGNMENTS

RESULT 1
B0516583
LOCUS
DEFINITION
B0516583
EST623998 Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues Solanum tuberosum cDNA clone STM1Y71 5' end, mRNA sequence.

ACCESSION
B0516583
VERSION
B0516583.1 GI:21375452
KEYWORDS
EST.
SOURCE
Solanum tuberosum (potato)
ORGANISM
Solanum tuberosum

REFERENCE
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamiales; Solanales; Solanaceae; Solanum.

AUTHORS
1 (bases 1 to 820)
Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C., Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J., and Karanaycheva, S.A.
Generation of a set of potato cDNA clones for microarray analyses
Unpublished (2002)
Other ESTs: EST623999
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@tigr.org
This clone can be obtained from the University of Arizona Genomics Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T3.

FEATURES
Location/Qualifiers
1..820

Source
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultiivar="Kennebec or Bintje"
/db_xref="taxon:4113"
/clone="STM1Y71"
/issue="STM1Y71"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/clone_id="Generation of a set of potato cDNA clones for microarray analyses mixed potato tissues"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; supplier: Combination of untreated and phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes, tubers, or roots."

ORIGIN

Query Match 64.3%; Score 816.8; DB 5; Length 820;
 Best Local Similarity 99.8%; Pred. No. 1.1e-214;
 Matches 818; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 320 CACGACAAATTCCTCAAGATGAGTGAAGTCTCTTTTCGGGAAGTGTTCACAGTTCTAGA 379
DB 1 CACGACAAATTCCTCAAGATGAGTGAAGTCTCTTTTCGGGAAGTGTTCACAGTTCTAGA 60
QY 380 AGAATGTTGAGCAATAATGCTGCTGCTGCTTCAATGAGAGTTACTAATCTGAGGT 439
DB 61 AGAATGTTGAGCAATAATGCTGCTGCTGCTTCAATGAGAGTTACTAATCTGAGGT 120
QY 440 TGACATCTTTGATAGAGAAAAGTATTCATTTGAAACGTTCTTAAGACGTTGTCAGAA 499
DB 121 TGACATCTTTGATAGAGAAAAGTATTCATTTGAAACGTTCTTAAGACGTTGTCAGAA 180
QY 500 ATTTCACAAATTGAAGTGTGATGAGCATGTTACCAACCATTTGATGCTTTAAGTTGT 559
DB 181 ATTTCACAAATTGAAGTGTGATGAGCATGTTACCAACCATTTGATGCTTTAAGTTGT 240
QY 560 TGAATCTTGCACTGAAGATTGTTGACAGCACTGCAACCCCAACATCTGTTTGA 619
DB 241 TGAATCTTGCACTGAAGATTGTTGACAGCACTGCAACCCCAACATCTGTTTGA 300
QY 620 CAGGAATCTCTCTTCCAAAGGGGCTTACACCCGATATTACTGCTTCCAGTCTCAA 679
DB 301 CAGGAATCTCTCTTCCAAAGGGGCTTACACCCGATATTACTGCTTCCAGTCTCAA 360
QY 680 AAGAGAGATCCACAGGAGGCACTGTGTCAAGCTGTACAGTGGAAAGTAAAGATTTT 739
DB 361 AAGAGAGATCCACAGGAGGCGCTGTGTGTCAAGCTGTACAGTGGAAAGTAAAGATTTT 420
QY 740 TCTTGGAAGTGAATGCTCTCATGATAGAAGAAAGAGTCTTGTGATGTGC 799
DB 421 TCTTGGAAGTGAATGCTCTCATGATAGAAGAAAGAGTCTTGTGATGTGC 480
QY 800 TGGATTTTACATGACCTGTAGCTGTGATGATGAGGAGGTTGTTGAAAAGAAAA 859
DB 481 TGGATTTTACATGACCTGTAGCTGTGATGATGAGGAGGTTGTTGAAAAGAAAA 540
QY 860 TGCATCGACAAAGCTTGAAGATTCAGTCACTCAATGAGCAAGATTTTATGGGCTGC 919
DB 541 TGCATCGACAAAGCTTGAAGATTCAGTCACTCAATGAGCAAGATTTTATGGGCTGC 600
QY 920 TAGGAACAATCAAGAATTAAAGTGAAGAGCCATGAGAGTACCCGAATCTTTTC 979
DB 601 TAGGAACAATCAAGAATTAAAGTGAAGAGCCATGAGAGTACCCGAATCTTTTC 660
QY 980 TTATGATCAAGAGATTTATTCATGCTGTTGCTGTTGAATGCTGATGCTGCGGC 1039
DB 661 TTATGATCAAGAGATTTATTCATGCTGTTGCTGTTGAATGCTGATGCTGCGGC 720
QY 1040 TCCCTCTGAGATCAATTTGCTATTTGATCTGTAATTTGATTTCAACCAAGATAT 1099
DB 721 TCCCTCTGAGATCAATTTGCTATTTGATCTGTAATTTGATTTCAACCAAGATAT 780
QY 1100 AGACTGTAGGTATGATCTTTTCTTTCATGTTGATTAGA 1139
DB 781 AGACTGTAGGTATGATCTTTTCTTTCATGTTGATTAGA 820

```

RESULT 2
 CK252266 1007 bp mRNA linear EST 30-JUL-2004
 LOCUS ESTY35903 potato callus cDNA library, normalized and full-length
 DEFINITION Solanum tuberosum cDNA clone POC8781 5' end, mRNA sequence.
 ACCESSION CK252266
 VERSION CK252266.1 GI:39806092
 KEYWORDS EST.
 SOURCE Solanum tuberosum (potato)
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE
 1 (bases 1 to 1007)
 AUTHORS Buell, C.R., Hart, A., Zismann, V., Karanycheva, S.A. and Baker, B.
 TITLE Generation of ESTs from potato callus tissue
 JOURNAL Unpublished (2003)
 COMMENT Contact: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-aray@igrr.org
 Clones can be requested from the University of Arizona Genomics
 Institute via <http://genome.arizona.edu/orders/>.
 Seq primer: ATG TAG GTG ACA CTA TAG.
 Location/Qualifiers

FEATURES
 source
 1..1007
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultiivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="POC8781"
 /tissue_type="callus"
 /lab_host="DH10B-Tona"
 /clone_lib="potato callus cDNA library, normalized and
 full-length"
 /note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
 supplier: RNA was isolated from Solanum tuberosum var.
 Kennebec callus tissue grown on solid media."

ORIGIN

Query Match 60.4%; Score 767.4; DB 7; Length 1007;
 Best Local Similarity 99.2%; Pred. No. 5.4e-201;
 Matches 771; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

QY 2 TGCAAAATGAGAGCTCTCAATACACACACCGATGATTGGCAGTCTTCCGTATG 61
DB 231 TGCAAAATGAGAGCTCTCAATACACACACCGATGATTGGCAGTCTTCCGTATG 290
QY 62 TGATGTTCTTAAAGCAATGTTGTTCTCAGAGTGCATCACTTTGGAGGCAATGCTAT 121
DB 291 TGATGTTCTTAAAGCAATGTTGTTCTCAGAGTGCATCACTTTGGAGGCAATGCTAT 350
QY 122 GCCAAATTTGAAGGCTCTCACTACCACTGCTGCGGTGATGACATACCGGAGCCGAT 181
DB 351 GCCAAATTTGAAGGCTCTCACTACCACTGCTGCGGTGATGACATACCGGAGCCGAT 410
QY 182 ATTGAATCTTTACCTGTTGATGATGATTTCAACCTCTTATGACATTTATTTGACAGA 241
DB 411 ATTGAATCTTTACCTGTTGATGATGATTTCAACCTCTTATGACATTTATTTGACAGA 470
QY 242 TACAACCAAGCTATGGAATCAAACTAGCAGAGAGCCAGTCTGTTGGGTTGA 301
DB 471 TACAACCAAGCTATGGAATCAAACTAGCAGAGAGCCAGTCTGTTGGGTTGA 530
QY 302 GTTGATCCCTGCTGTGTCAGACAAATTTCTCAAGATGAGTGAATCTTTTGGGAA 361
DB 531 GTTGATCCCTGCTGTGTCAGACAAATTTCTCAAGATGAGTGAATCTTTTGGGAA 590
QY 362 GTTGTTACCAAGTTCTTCAAGAAATGTTGAGATTAATGCTGCTGTGTTCAATGAGA 421
DB 591 GTTGTTACCAAGTTCTTCAAGAAATGTTGAGATTAATGCTGCTGTGTTCAATGAGA 650
QY 422 GGTTACTAATCTGAGGTTGATGATGATGAGAAAGGATTTCAATGAAAAGGTTCT 481
DB 651 GGTTACTAATCTGAGGTTGATGATGATGAGAAAGGATTTCAATGAAAAGGTTCT 710
QY 482 AAGACCGTTGGGAGAAATTTCCACAAATTGAAGTGTGATGAGAGATGTTACACCAT 541
DB 711 AAGACCGTTGGGAGAAATTTCCACAAATTGAAGTGTGATGAGAGATGTTACACCAT 770
QY 542 TGATGCTGTAAATTTGTTGATCTTGCATGGAAGATTTGTTGGACACATGTCACCCC 601
DB 771 TGATGCTGTAAATTTGTTGATCTTGCATGGAAGATTTGTTGGACACATGTCACCCC 830
QY 602 ACAACATCTGTTTGAACAGAAATTCCTCTTCCAAAGGGGCTTACACCGCATATTATTA 661

```

Db 831 ACACATCTTGTGTAACGAGATTTCTCTTCAAGGGGGCTTACACCGCATATTA 890
Qy 662 CTGCCTTCAGTCTCAAAAAGAGATCCACAGGAGGCACTTGTCACCTGTAACAAG 721
Db 891 CTGCCTTCAGTCTCAAAAAGAGATCCACAGGAGGCGCTTGTCAGCTGTAACAAG 950
Qy 722 TGGAGTAAAGATTTTCTTGGAGCTGATAGTCTCTCATGATAGACGAAGAA 778
Db 951 TGGAGTAAAGATTTTCTTGGAGCTGATAGTCTCTCATGATAGACGAAGAA 1007

RESULT 3
B0516584/c 751 bp mRNA linear EST 07-MAR-2003
LOCUS EST623999 Generation of a set of potato cDNA clones for microarray
DEFINITION analyses mixed potato tissues Solanum tuberosum cDNA clone STM1Y71
3' end, mRNA sequence.
B0516584
B0516584
B0516584.1 GI:21375453
VERSION
KEYWORDS
SOURCE
ORGANISM Solanum tuberosum (potato)
Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
1 (bases 1 to 751)
Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,
Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and
Karamecheva, S.A.
Generation of a set of potato cDNA clones for microarray analyses
Unpublished (2002)
Other ESTs: EST623998
COMMENT
TITLE The Institute for Genomic Research
JOURNAL 9712 Medical Center Dr, Rockville, MD 20850, USA
EMAIL: potato-array@ligr.org
This clone can be obtained from the University of Arizona Genomics
Institute. Orders can be made through URL:
http://genome.arizona.edu/orders/
Seq primer: T7
FEATURES
Location/Qualifiers
source
1..751
/organism="Solanum tuberosum"
/mol_type="mRNA"
/culturvar="Kennebec or Binje"
/db_xref="taxon:4113"
/clone="STM1Y71"
/issue_type="mixed tissues"
/lab_host="SOLR"
/clone_idb="Generation of a set of potato cDNA clones for
microarray analyses mixed potato tissues"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: combination of untreated and phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants; petioles, germinating
eyes, tubers, or roots."

Qy 438 GTTGACATGTTGTAGAGAAAGATATTCAATTGAAACGTTCTAAGACCGTTGGTCAG 497
Db 751 GGTGACATGTTGTAGAGAAAGATATTCAATTGAAACGTTCTAAGACCGTTGGTCAG 692
Qy 498 AAATTTCCACATTAAGAGTCGTGATGAGACATGTTACACCATTTGATGCTGTTAAGTT 557
Db 691 AAATTTCCACATTAAGAGTCGTGATGAGACATGTTACACCATTTGATGCTGTTAAGTT 632
Qy 558 GTTGAATCTTGCACTGAAGATTTGTCAGAAATGACACCCCAACATCTTGTGTTG 617

Db 631 GTTGAATCTTGCACTGAAGATTTGTCAGAACTGTACCCCAACATCTTGTGTTG 572
Qy 618 AACAGAAATTTCTCTTCCAAAGGGGCTTACAACCGCATATTTACGCTTCAGTCTC 677
Db 571 AACAGAAATTTCTCTTCCAAAGGGGCTTACAACCGCATATTTACGCTTCAGTCTC 512
Qy 678 AAAAGAGATCCACAGGAGGCACTTGTCAGCTGTAAACAAGTGAAGTAAAGATTT 737
Db 511 AAAAGAGATCCACAGGAGGCGCTTGTCAGCTGTAAACAAGTGAAGTAAAGATTT 452
Qy 738 TTTCTTGGAGCTGATAGTCTCTCTCATGATAGACGAAGAAAGAGTGTGATGT 797
Db 451 TTTCTTGGAGCTGATAGTCTCTCTCATGATAGACGAAGAAAGAGTGTGATGT 392
Qy 798 GCTGATTTTACAAAGCAGCTGAGCTGTACGATATGCGAAGGTGTTGAAAAGAA 857
Db 391 GCTGATTTTACAAAGCAGCTGAGCTGTACGATATGCGAAGGTGTTGAAAAGAA 332
Qy 858 AATGCACTGACAAAGCTTGAAGCATTCAGTACCTTCAATGACCAAGATTTTATGAGCTT 917
Db 331 AATGCACTGACAAAGCTTGAAGCATTCAGTACCTTCAATGACCAAGATTTTATGAGCTT 272
Qy 918 CCTAGAACAACTCAAAAGATTAAGTTAGTAGAGCCATGAGAGGTACCCGAATCTTT 977
Db 271 CCTAGAACAACTCAAAAGATTAAGTTAGTAGAGCCATGAGAGGTACCCGAATCTTT 212
Qy 978 TCTTATGACATCAGAAATATTTATCCAGTGTGCTGTAATGCTCGAGTGGTCCG 1037
Db 211 TCTTATGACATCAGAAATATTTATCCAGTGTGCTGTAATGCTCGAGTGGTCCG 152
Qy 1038 GCTCCTCTGAGAAATCATTTGATCTTCTTCAATGATTAATGATTCACCAAGAT 1097
Db 151 GCTCCTCTGAGAAATCATTTGATCTTCTTCAATGATTAATGATTCACCAAGAT 92
Qy 1098 ATAGACTGATGATGATATCTTTCTTCAATGATTAATGATTCACCAAGAT 1157
Db 91 ATAGACTGATGATGATATCTTTCTTCAATGATTAATGATTCACCAAGAT 32
Qy 1158 ATCCTTTCAGCTTAATTAATTAATGAAACAT 1188
Db 31 ATCCTTTCAGCTTAATTAATTAATGAAACAT 1

RESULT 4
CK253198 915 bp mRNA linear EST 30-JUL-2004
LOCUS EST736835 potato callus cDNA library, normalized and full-length
DEFINITION Solanum tuberosum cDNA clone POC351 5' end, mRNA sequence.
CK253198
VERSION CK253198.1 GI:39807898
KEYWORDS
SOURCE
ORGANISM Solanum tuberosum (potato)
Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
1 (bases 1 to 915)
Buell, C.R., Hart, A., Zisemann, V., Karamecheva, S.A. and Baker, B.
Generation of ESTs from potato callus tissue
Unpublished (2003)
Other ESTs: EST736836
COMMENT
TITLE The Institute for Genomic Research
JOURNAL 9712 Medical Center Dr, Rockville, MD 20850, USA
EMAIL: potato-array@ligr.org
Clones can be requested from the University of Arizona Genomics
Institute via http://genome.arizona.edu/orders/
Seq primer: ATT TAG GTG ACA CTA TAG.
FEATURES
Location/Qualifiers
source
1..915
/organism="Solanum tuberosum"
/mol_type="mRNA"
/culturvar="Kennebec"

```

/db xref="taxon:4113"
/clone="POCC351"
/tissue_type="callus"
/lab_host="DH10B-Tona"
/clone_lib="potato callus cDNA library, normalized and
full-length"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRI, Site 2: NotI;
supplier: RNA was isolated from Solanum tuberosum var.
Kennebec callus tissue grown on solid media."

```

ORIGIN

Query Match

```

Best Local Similarity 54.4%; Score 690.8; DB 7; Length 915;
Matches 695; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 2 TGCAGAAATGAGGCTCTCAATCAACACAGTGAATGGCATCTTCCGTGATG 61
Db 214 TGCAGAAATGAGGCTCTCAATCAACAGTGAATGGCATCTTCCGTGATG 273
OY 62 TGAATGTTTAAAGGAGTGTCTCTCAAGTGCATCACTTTGGAGGGCAATGAT 121
Db 274 TGAATGTTTAAAGGAGTGTCTCTCAAGTGCATCACTTTGGAGGGCAATGAT 333
OY 122 GCCAAATTTGAAGCCCTTATCACTACCTGCTGCTGTAACATACCGGAGCGAT 181
Db 334 GCCAAATTTGAAGCCCTTATCACTACCTGCTGCTGTAACATACCGGAGCGAT 393
OY 182 ATTGAATCTTTAAGTGTGATGATTTCAACCTCTTAAGACATTTATTGACAGA 241
Db 394 ATTGAATCTTTAAGTGTGATGATTTCAACCTCTTAAGACATTTATTGACAGA 453
OY 242 TACAAACAGTCTCATGGAATCAAACTGCAAGAGAGCCAGTCTGATTGGGGTGA 301
Db 454 TACAAACAGTCTCATGGAATCAAACTGCAAGAGAGCCAGTCTGATTGGGGTGA 513
OY 302 GTTGTACCTCTGCTGTCGACGACAAATTTCTCAAGATGAGTGACTGATCTTT 361
Db 514 GTTGTACCTCTGCTGTCGACGACAAATTTCTCAAGATGAGTGACTGATCTTT 573
OY 362 GTGTTTACCAAGTTCTCAAGAAATGTTGAGCATTAATGCTGCTGTTATGAGA 421
Db 574 GTGTTTACCAAGTTCTCAAGAAATGTTGAGCATTAATGCTGCTGTTATGAGA 633
OY 422 GTTATCTAATCTGAGGTTGACATGTTGATGAGAAAGTATTCATTGAACGGTCT 481
Db 634 GTTATCTAATCTGAGGTTGACATGTTGATGAGAAAGTATTCATTGAACGGTCT 693
OY 482 AAGACCGTTGAGAGAAATTTCCAAATTTGAAGGCTGTGATGAGACATGTTACCA 541
Db 694 AAGACCGTTGAGAGAAATTTCCAAATTTGAAGGCTGTGATGAGACATGTTACCA 753
OY 542 TGATGCTGTTAAGTTGTTGATCTTGCATGAAAGATTGTTGACGAACTGTACCC 601
Db 754 TGATGCTGTTAAGTTGTTGATCTTGCATGAAAGATTGTTGACGAACTGTACCC 813
OY 602 ACAACATCTGTTTGAAGCAATTTCTCTTCCAAAGGGCTTACACCCCATATTA 661
Db 814 ACAACATCTGTTTGAAGCAATTTCTCTTCCAAAGGGCTTACACCCCATATTA 873
OY 662 CTGCTTTCAGTCTCTCAAGAGATCCACAGGAGGAGCACT 703
Db 874 CTGCTTTCAGTCTCTCAAGAGATCCACAGGAGGAGCACT 915

```

RESULT 5

CK242513/c

LOCUS

EST726150 potato callus cDNA library, normalized and full-length

DEFINITION

CK242513

ACCESSION

VERSION

KEYWORDS

SOURCE

Solanum tuberosum (potato)

ORGANISM

Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
1 (bases 1 to 790)
Buell, C.R., Hart, A., Zismann, V., Karamycheva, S.A. and Baker, B.
Generation of ESTs from potato callus tissue
Unpublished (2003)
Other ESTs: EST726147 EST726148 EST726149
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@igr.org
Institute via http://genome.arizona.edu/orders/
Seq primer: GTA ATA CGA CTC ACT ATA GGG C.
Location/Qualifiers
1..790

COMMENT

FEATURES

source

ORIGIN

```

Query Match 54.3%; Score 689.6; DB 7; Length 790;
Best Local Similarity 95.9%; Pred. No. 1.8e-179;
Matches 727; Conservative 0; Mismatches 9; Indels 22; Gaps 1;

```

```

OY 497 GAAATTTCCAAATTTGAAGTCTGATGACATGTTACCATTTGATGCTTTAAGTT 556
Db 790 GAAATTTCCAAATTTGAAGTCTGATGACATGTTACCATTTGATGCTTTAAGTT 731
OY 557 TGTGAATCTTTCAGTGAAGATTTGTCAGCAATGTCACCCCAACATCTTTT 616
Db 730 TGTGAATCTTTCAGTGAAGATTTGTCAGCAATGTCACCCCAACATCTTTT 671
OY 617 GAAAGGAATTCCTCTTCCAAAGGGCTTACACCGCATTAATCTGCTTCCAGTCT 676
Db 670 GAAAGGAATTCCTCTTCCAAAGGGCTTACACCGCATTAATCTGCTTCCAGTCT 611
OY 677 CAAAAGAGATCCACAGGAGGCACTTGTGACGCTGTAACAGTGAAGTAAAGATT 736
Db 610 CAAAAGAGATCCACAGGAGGCGCTTGTGACGCTGTAACAGTGAAGTAAAGATT 551
OY 737 TTTTCTGGAGCTGATGATGCTCTCTCATGATGACGAAAGAAAGTCTTGTGATG 796
Db 550 TTTTCTGGAGCTGATGATGCTCTCTCATGATGACGAAAGAAAGTCTTGTGATG 491
OY 797 TGTGTAATTTTCAATGACCTGTAGCCTTGTGAGTAAATGCGAAGCTTTGAAA 856
Db 490 TGTGTAATTTTCAATGACCTGTAGCCTTGTGAGTAAATGCGAAGCTTTGAAA 431
OY 857 AAATGATCTCGCAAGCTTGAAGCATTTCACTAGCTTCAATGACCAATTTTATG 916
Db 430 AAATGATCTCGCAAGCTTGAAGCATTTCACTAGCTTCAATGACCAATTTTATG 371
OY 917 TCTTAGAACAACCTCAAGATTAGTTAGTAAGCCCTGGAAGTACCCGAATCTT 976
Db 370 TCTTAGAACAACCTCAAGATTAGTTAGTAAGCCCTGGAAGTACCCGAATCTT 311
OY 977 TTCTTATGATCAGAGATATTTATCCATGTTGCTGTGAATGCTCGATGTTGCC 1036
Db 310 TTCTTATGATCAGAGATATTTATCCATGTTGCTGTGAATGCTCGATGTTGCC 251
OY 1037 GGCTCCTCTTGAAGA-----TCATTGTCAATTTCTTGAATGCT 1074

```

Db 250 GGGTCTCTCTGAGAAAGGAAAAAGAGATTCCTCTTCATTTGCAATCTTTGACGT 191
QY 1075 AATATGTGATTAACCAAGATATAGATAGTATGATATCATCTTTCTTTGATGGA 1134
Db 190 AATATGTGATTAACCAAGATATAGATAGTATGATATCATCTTTCTTTGATGGA 131
QY 1135 TTAGATATTATACGATGATATATCTTTGACGTAATAAATTATGGAACATTAAGCTT 1194
Db 130 TTGATATTATACCGATGATATATCTTTGACGTAATAAATTATGGAACATTAAGCTT 71
QY 1195 TGCACGCTCACCAAGGCTCTCTGATATCTTGAAGTCT 1232
Db 70 TGCACGCTCACCAAGGCTCTCTGATATCTTGAAGATTT 33
RESULT 6 920 bp mRNA linear EST 30-JUL-2004
LOCUS CK250612
DEFINITION EST734249 potato callus cDNA library, normalized and full-length
Solanium tuberosum CDNA clone POCBH29 5' end, mRNA sequence.
ACCESSION CK250612 GI:39802848
VERSION CK250612.1
KEYWORDS EST.
SOURCE Solanium tuberosum (potato)
ORGANISM Solanium tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanium.
REFERENCE 1 (bases 1 to 920)
AUTHORS Buell, C.R., Hart, A., Zisemann, V., Karamycheva, S.A. and Baker, B.
TITLE Generation of ESTs from potato callus tissue
JOURNAL Unpublished (2003)
COMMENT Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@igr.org
Clones can be requested from the University of Arizona Genomics
Institute via http://genome.arizona.edu/orders/
Seq primer: ATT TAG GTC ACA CTA TAG.
FEATURES
source
1. .920
/organism="Solanium tuberosum"
/mol_type="mRNA"
/cultiyar="Kennebec"
/db_xref="taxon:4113"
/clone="POCBH29"
/issue_type="callus"
/lab_host="DH10B-Tona"
/clone_lib="potato callus cDNA library, normalized and
full-length"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: RNA was isolated from Solanium tuberosum var.
Kennebec callus tissue grown on solid media."

ORIGIN
Query Match 52.1% Score 662.6; DB 7; Length 920;
Best Local Similarity 99.3%; Pred. No. 5.6e-172; Indels 1; Gaps 1;
Matches 676; Conservative 0; Mismatches 4;
2 TGC AAAAATGAGCTCTCAATCACACAACTGATGATGGCATCTTCCGCTGATGG 61
Db 241 TGC AAAAATGAGCTCTCAATCACACAACTGATGATGGCATCTTCCGCTGATGG 300
QY 62 TGATGTTCTTAAGGAGTGTCTCTCACAGTGCACATCACTTTGGAGGCAATGTGAT 121
Db 301 TGATGTTCTTAAGGAGTGTCTCTCACAGTGCACATCACTTTGGAGGCAATGTGAT 360
QY 122 GCCAATTTGAAGCTCTCATACATACAGTGTCTGCTGTAGCATACCGGAGGCGAT 181
Db 361 GCCAATTTGAAGCTCTCATACATACAGTGTCTGCTGTAGCATACCGGAGGCGAT 420
QY 162 ATTGAATCTTTACTGTTGATGATTTCAACCTCTTATGACACTTTATTTGACAGA 241

Db 421 ATTGAATCTTTACTGTTGATGATTTCAACCTCTTATGACACTTTATTTGACAGA 480
QY 242 TACAAACCACTCTTATGGAATATCAACTACGAGAGAGACCAGTGTATTTGGGGTGA 301
Db 481 TACAAACCACTCTTATGGAATATCAACTACGAGAGAGACCAGTGTATTTGGGGTGA 540
QY 302 GTTGTACCTGCTGGTGGCCAGCAAAATCTCAAGATGAGAGTATCTTTCCGGGA 361
Db 541 GTTGTACCTGCTGGTGGCCAGCAAAATCTCAAGATGAGAGTATCTTTCCGGGA 600
QY 362 GTTGTACCACTCTTACCAAGAAATGTTGAGCATTAATATGCTTGTGTTGATGAG 421
Db 601 GTTGTACCACTCTTACCAAGAAATGTTGAGCATTAATATGCTTGTGTTGATGAG 660
QY 422 GGTACTTAATCTGAGGTGACATGTTGATGAGAAAGGTTTATTTGAAACGGTCTT 481
Db 661 GGTACTTAATCTGAGGTGACATGTTGATGAGAAAGGTTTATTTGAAACGGTCTT 720
QY 482 AAGACCGTTGGTGCAGAAATTTCCAGATTTGAAGGTCGTATGAGCATGTTCACCAT 541
Db 721 AAGACCGTTGGTGCAGAAATTTCCAGATTTGAAGGTCGTATGAGCATGTTCACCAT 780
QY 542 TGATGCTGTTAAGTTGTTGAATCTTGCACTGAGAGATTTGTTGCAGCACTGTACCCC 601
Db 781 TGATGCTGTTAAGTTGTTGAATCTTGCACTGAGAGATTTGTTGCAGCACTGTACCCC 840
QY 602 ACAACATCTTTGTTGAACGAAATCTCTTTCAGAGGGGCTTACACCGCATATTA 661
Db 841 ACAACATCTTTGTTGAACGAAATCTCTTTCAGAGGGGCTTAC-ACCGCATATTA 899
QY 662 CTGCTTCCAGTCTCAAAAG 682
Db 900 CTGCTTCCAGTCTCAAAAG 920

RESULT 7
CK249779/c 825 bp mRNA linear EST 30-JUL-2004
LOCUS EST734346 potato callus cDNA library, normalized and full-length
DEFINITION Solanium tuberosum CDNA clone POCBA88 3' end, mRNA sequence.
ACCESSION CK249779
VERSION CK249779.1 GI:39801235
KEYWORDS EST.
SOURCE Solanium tuberosum (potato)
ORGANISM Solanium tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanium.
REFERENCE 1 (bases 1 to 825)
AUTHORS Buell, C.R., Hart, A., Zisemann, V., Karamycheva, S.A. and Baker, B.
TITLE Generation of ESTs from potato callus tissue
JOURNAL Unpublished (2003)
COMMENT Other ESTs: EST733415
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@igr.org
Clones can be requested from the University of Arizona Genomics
Institute via http://genome.arizona.edu/orders/
Seq primer: GTA ATA CGA CTC ACT ATA GGG C.
FEATURES
source
1. .825
/organism="Solanium tuberosum"
/mol_type="mRNA"
/cultiyar="Kennebec"
/db_xref="taxon:4113"
/clone="POCBA88"
/issue_type="callus"
/lab_host="DH10B-Tona"
/clone_lib="potato callus cDNA library, normalized and
full-length"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;

QY 1142 TTATCAGATGATATATCTTACGTAATAATATGAAACAATAGCTTGGACGC 1201
| | | | |
DB 123 TTATCCGATGATAAATCTTTACGTAATAATATGAAACAATAGCTTGGACGC 64
| | | | |
QY 1202 TCACCAAGTCTCTGATATTTGAACTTCT 1232
| | | | |
DB 63 TCACCAAGTCTCTGATATTTGAACTTCT 33
| | | | |
RESULT 9 708 bp mRNA linear EST 30-JULY-2004
CK251510/c
LOCUS EST375147 potato callus cDNA library, normalized and full-length
DEFINITION Solanum tuberosum cDNA clone POCB96 3' end, mRNA sequence.
ACCESSION CK251510
VERSION CK251510.1 GI:39804606
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
REFERENCE Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE asterids; lamids; Solanales; Solanaceae; Solanum.
JOURNAL 1 (bases 1 to 708)
COMMENT Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr. Rockville, MD 20850, USA
Email: potato-array@tigr.org
Clones can be requested from the University of Arizona Genomics
Institute via <http://genome.arizona.edu/orders/>.
Seq primer: GTA ATA CGA CTC ACT ATA GGC C.

FEATURES

source
1. 708
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="POCB96"
/tissue_type="callus"
/lab_host="DH10B-Tona"
/clone_id="potato callus cDNA library, normalized and
full-length"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: RNA was isolated from Solanum tuberosum var.
Kennebec callus tissue grown on solid media."

ORIGIN

Query Match 47.8%; Score 607.6; DB 7; Length 708;
Best Local Similarity 95.4%; Pred. No. 8.9e-157;
Matches 645; Conservative 0; Mismatches 9; Indels 22; Gaps 1;
QY 579 TTGTGTGAGCACTGTGACCCACACATCTTTGTAACAGGAATTTCTCTCCAA 638
| | | | |
DB 708 TTGTGTGAGCGAGTGTACCCACACATCTTTGTAACAGGAATTTCTCTCCAA 649
| | | | |
QY 639 GGGGGCTTACAAACGCAATTAATCTGCTTCCAGTCTCAAAAGAGATCCACAGGAG 698
| | | | |
DB 648 GGGGGCTTACAAACGCAATTAATCTGCTTCCAGTCTCAAAAGAGATCCACAGGAG 589
| | | | |
QY 699 GCACTTGTGAGCTGTGTAACAAGTGAAGTAAGATTTTCTTGGAGTGTAGTCT 758
| | | | |
DB 588 GCGCTTGTGAGCTGTGTAACAAGTGAAGTAAGATTTTCTTGGAGTGTAGTCT 529
| | | | |
QY 759 CCTCATGATTAACAAGAAAGAGTCTTCTGTGATGTGCTGTAATTTCAATGCACT 818
| | | | |
DB 528 CCTCATGATTAACAAGAAAGAGTCTTCTGTGATGTGCTGTAATTTCAATGCACT 469
| | | | |
QY 819 GTAGCTTGTGATTAACAAGAAAGTGTGTAAGAAAGAAATGCACTGCAAGCTTGA 878
| | | | |
DB 468 GTAGCTTGTGATTAACAAGAAAGTGTGTAAGAAAGAAATGCACTGCAAGCTTGA 409
| | | | |

QY 879 GCATTCAGTCTGATGATGACCAAGATTTTATAGGCTTCTAGGAACAACCTCAAGATT 938
| | | | |
DB 408 GCATTTACTAGCTTCAATGAGCAAGACTTTATAGGCTTCTAGGAACAACCTCAAGATT 349
| | | | |
QY 939 AAGTTAGTAAGACCGCATGGAAGGTACCCGAATCTTTTCTTATGATCAGAGATATT 998
| | | | |
DB 348 AAGTTAGTAAGACCGCATGGAAGGTACCCGAATCTTTTCTTATGATCAGAGATATT 289
| | | | |
QY 999 ATTCCCATTTTCTGCTGTAAGTAAGTGTGAGCTGCTGCTGAGAA----- 1052
| | | | |
DB 288 ATTCCCATTTTCTGCTGTAAGTAAGTGTGAGCTGCTGCTGCTGAGAAAGGAA 229
| | | | |
QY 1053 -----TCATTTGTCACTTGTGATGTAATATGATGATCAACCAAGA 1096
| | | | |
DB 228 AAGAGATTCTCTTCTCAATTTGTCTTCTGACGTGAATATGATGATCAACCAAGA 169
| | | | |
QY 1097 TATAGACTGAGTGTATGATCATCTTTTCTTATGATGATGATGATGATGATTA 1156
| | | | |
DB 168 TATAGACTGAGTGTATGATCATCTTTTCTTATGATGATGATGATGATGATTA 109
| | | | |
QY 1157 TATCTTTTCACTTAATTAATTAATGAAACAATAAGCTTGGACAGCTCACCAAGTCTCC 1216
| | | | |
DB 108 TATCTTTTCACTTAATTAATTAATGAAACAATAAGCTTGGACAGCTCACCAAGTCTCC 49
| | | | |
QY 1217 TGTATCTGAACTTCT 1232
| | | | |
DB 48 TGTATCTGAACTTCT 33
| | | | |

RESULT 10

BI927946 732 bp mRNA linear EST 18-OCT-2001
LOCUS EST547823 tomato flower, 3 - 8 mm buds Lycopersicon esculentum cDNA
DEFINITION clone cTOB24A1 5' end, mRNA sequence.
ACCESSION BI927946
VERSION BI927946.1 GI:16238481
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
REFERENCE Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
TITLE asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
JOURNAL 1 (bases 1 to 732)
COMMENT van der Hoeven, R.S., Bezzerides, J.L., Karmycheva, S.A., Tsai, J.,
Utechtback, T., Van Aken, S., Koning, C.M., Nierman, W., Fraser, C.M.,
Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.
Generation of ESTs from tomato flower tissue, 3-8 mm buds (2001)
Contact: CUCI
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
This clone is available through the Clemson University Genomics
Institute
Seq primer: T3.

FEATURES

source
1. 732
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cTOB24A1"
/tissue_type="flower"
/dev_stage="3-8mm buds"
/clone_id="tomato flower, 3 - 8 mm buds"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Cornell University; sequencing: The
Institute for Genomic Research. Flower buds and flowers
were taken from greenhouse plants (4-6 wks old, T496).
They were immediately frozen in liquid nitrogen and then
size-separated while remaining frozen."

ORIGIN

Query Match 47.2%; Score 599.4; DB 4; Length 732;
Best Local Similarity 97.4%; Pred. No. 1.7e-154;
Matches 609; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```
QY 2 TGCAGAAATGAGAGCTCTCAATCAACACACCTGATGATGGCATCTTCATCTCCGTGATGG 61
DB 108 TGCAGAAATGAGAGCTCTCAATCAACACACCTGATGATGGCATCTTCATCTCCGTGATGG 167
QY 62 TGAATGCTTAAGGAGGAGTGTCTCTCAAGTGAACATCACTTTGGAGGGCAATATGCTAT 121
DB 168 TGAATGCTTAAGGAGGAGTGTCTCTCAAGTGAACATCACTTTGGAGGGCAATATGCTAT 227
QY 122 GCCAATTTGAAGCCTCTCAATCACTCACTGCTGCTGCTGATGACATACCGGAGGCGAT 181
DB 228 GCCAATTTGAAGCCTCTCAATCACTCACTGCTGCTGCTGATGACATACCGGAGGCGAT 287
QY 182 ATGAATCTTTAAGCCTGATGATGATGATTTCAACCTCTTATGACACTTTATTTGACAGA 241
DB 288 ATGAATCTTTAAGCCTGATGATGATGATTTCAACCTCTTATGACACTTTATTTGACAGA 347
QY 242 TACAACCAAGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
DB 348 TACAACCAAGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 407
QY 302 GTTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361
DB 408 GTTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 467
QY 362 GTTATACCAAGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 421
DB 468 GTTATACCAAGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 527
QY 422 GGTATCTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 481
DB 528 GGTATCTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 587
QY 482 AAGACCGTGTGTCAGAAATTTCCACAATTGAAGTGTGATGATGATGATGATGATGATG 541
DB 588 AAGACCGTGTGTCAGAAATTTCCACAATTGAAGTGTGATGATGATGATGATGATGATG 647
QY 542 TGATGCTTTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 601
DB 648 TGATGCTTTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 707
QY 602 ACAACATCTTGTGTTGAACAGGAT 626
DB 708 ACAACATCTTGTGTTGAACAGGAT 732
```

RESULT 11
CK242511 957 bp mRNA linear EST 30-UTL-2004
LOCUS EST726148 potato callus cDNA library, normalized and full-length
DEFINITION Solanum tuberosum cDNA clone POCA005 5' end, mRNA sequence.
ACCESSION CK242511
VERSION CK242511.1 GI:39786207
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Bukayocra; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.
1 (bases 1 to 957)
Buell, C.R., Hart, A., Ziemann, V., Karameycheva, S.A. and Baker, B.
Generation of ESTs from potato callus tissue
Unpublished (2003)
Other ESTs: EST726147 EST726149 EST726150
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato-array@igr.org

Clones can be requested from the University of Arizona Genomics

Institute via <http://genome.arizona.edu/orders/>
Seq primer: ATT TAG GTG ACA CTA TAG.
Location/Qualifiers

FEATURES

SOURCE

1. .957

/organism="Solanum tuberosum"

/mol_type="mRNA"

/cultivar="Kennebec"

/db_xref="taxon:4113"

/clone="POCA005"

/issue_type="callus"

/lab_host="DH10B-Tona"

/clone_lib="potato callus cDNA library, normalized and"

/note="Vector: PCWVspore6.1; Site 1: EcoRI; Site 2: NotI; supplier: RNA was isolated from Solanum tuberosum var. Kennebec callus tissue grown on solid media."

ORIGIN

Query Match 46.8%; Score 595; DB 7; Length 957;
Best Local Similarity 99.0%; Pred. No. 3e-153;
Matches 609; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

```
QY 2 TGCAGAAATGAGAGCTCTCAATCAACACACCTGATGATGGCATCTTCATCTCCGTGATGG 61
DB 341 TGCAGAAATGAGAGCTCTCAATCAACACACCTGATGATGGCATCTTCATCTCCGTGATGG 400
QY 62 TGAATGCTTAAGGAGGAGTGTCTCTCAAGTGAACATCACTTTGGAGGGCAATATGCTAT 121
DB 401 TGAATGCTTAAGGAGGAGTGTCTCTCAAGTGAACATCACTTTGGAGGGCAATATGCTAT 460
QY 122 GCCAATTTGAAGCCTCTCAATCACTCACTGCTGCTGCTGATGACATACCGGAGGCGAT 181
DB 461 GCCAATTTGAAGCCTCTCAATCACTCACTGCTGCTGCTGATGACATACCGGAGGCGAT 520
QY 182 ATGAATCTTTAAGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 241
DB 521 ATGAATCTTTAAGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 580
QY 242 TACAACCAAGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
DB 581 TACAACCAAGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 640
QY 542 TGATGCTTTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 601
DB 641 TGATGCTTTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 700
QY 602 ACAACATCTTGTGTTGAACAGGAT 626
DB 701 ACAACATCTTGTGTTGAACAGGAT 732
QY 422 GGTATCTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 481
DB 761 GGTATCTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 820
QY 482 AAGACCGTGTGTCAGAAATTTCCACAATTGAAGTGTGATGATGATGATGATGATGATG 541
DB 821 AAGACCGTGTGTCAGAAATTTCCACAATTGAAGTGTGATGATGATGATGATGATGATG 880
QY 542 TGATGCTTTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 601
DB 881 TGATGCTTTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 939
QY 602 ACAACATCTTGTGTT 616
DB 940 ACAACATCTTGTGTT 954
```

RESULT 12

CNS0A31G 1301 bp mRNA linear HTC 06-FEB-2004
LOCUS CNS0A31G
DEFINITION Arabidopsis thaliana Full-length cDNA Complete sequence from clone
GSLPBG902805 of Hormone Treated Callus of strain col-0 of
Arabidopsis thaliana (thale cress) ..

ACCESSION BX828455
VERSION BX828455.1 GI:42460819
KEYWORDS HTC; GSLT cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 1301)
REFERENCE Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C.,
Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome
Annotation
Unpublished
2 (bases 1 to 1301)
JOURNAL Genoscope.
REFERENCE Direct Submission
AUTHORS Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out
full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castelli
V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C.,
Schachter V., Weissenbach J., Salanoubat M.
URGV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis
genome released by MIPS (Munich Information center for Protein
Sequences) . 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
length
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
FEATURES
source
Location/Qualifiers
1..1301
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="COL-0"
/db_xref="taxon:3702"
/clone="GSLTRH902B05"
/issue_type="Hormone Treated Callus"
/plasmid="PCMVSPORT_6"
1..1301
/gene="At4g22930"
ORIGIN
Query Match 46.1%; Score 585.4; DB 3; Length 1301;
Best Local Similarity 72.9%; Pred. NO. 1.5e-150;
Matches 754; Conservative 0; Mismatches 281; Indels 0; Gaps 0;
3 GCAGAAATGAGGCTCTCAATACACACCAACCTGATGATGGGACTCTTCATCTCCGATGATGT 62
187 GTAAAGATGAGGCTCAACATACATCACTACCTATGATGATGGGACTCTTCATCTCCGATGATGT 246
63 GATGTTCTTAAGGAGTGTCTCTCACAGTGCACATCACTTTGGAGGGCAATAGTCATG 122
247 GATCTTCTTCATGCTGTGTGTTCCCAACAGTGCAGATATTTTAAGAGGCAATGTGATG 306
123 CCAAAATTTGAAGCTCTCTATACATCACTGCTGCTGTGATGATACATACCGGAGCGCAT 182
307 CCGAATCGAAGCTCTCTGTCACCTCCACTGCACTGCACTTATTAACCGAAATTTATC 366
183 TTGAATCTTAACTGTTGATGATGATTTAAACCTCTTATGACACTTATTTAGACAT 242
367 ATGAAGCTTTGCCATCTGAGAGTACTTTGATTCATTTATGACCTTTATTTATGACTGAC 426
243 ACAACCAAGTCTTATGAAATCAAACTACAGAGAGAGACCGATCTGATTTGGGGTGAAG 302
427 AAAAATCTACCTGAAGATCAAGGTTTCAAGGGAAGTGGTGTGTTATGCGGTGAAG 486
303 TTGTAACCTGCTGTGCCACGACAAATTTCTCAAGATGAGTGAATGATCTTTTGGGAAG 362

DB 487 CTGTACCTCCCGAGGACAAACCAACTCTCAAGATGATGTCACGATCTCTTTGGAAA 546
363 TGTTTACCAAGTTCTACAGAAATGTTGAGCATATATGCTCTGCTGTTATGAGAG 422
547 TGTCTTACCGGATCTAGAAAGATGTGTCAACAAATATGCTTGTGTTATGAGGAG 606
423 GTTACATATCTGAGGTTACATGTTGATGAGAAAGGATTTACTTGAACGTTCTTA 482
607 GTACAGATCCAGATATCAATGTTTGAACCGGAGAAATCTTCAITGAGAGATTTCT 666
483 AGACCTTGATGAGAAATTTCCAAATTTGAAGTGTGATGAGAGATGTTACACCAT 542
667 CAGCTCTATCCAAAGCGCTCCGACGTGAAGGTGATGAGACATCAACACATG 726
543 GATGCTGTTAAGTTGTTGATTTGATTTGATCTGATGAGAGATTTGTCACCACTGTCA 602
727 GATGCTGTAATTTGTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 786
603 CAAATCTGTTTGTGAAGGAATTTCTCTTCCAAAGGGGCTTACAAACCGCATTAATTAC 662
787 CAACATCTCTTCTCAACAGAAAGCGCTTTTCCAAAGTGTGATTTACAACTCAACTAC 846
663 TGCTTCCAGTCTCTCAAAAGAGATCCACAGGAGGCACTTGTCACTGATTAACAGT 722
847 TGCTTCCAGTCTCTCAAAAGAGATCCACAGGAGGCACTTGTCACTGATTAACAGT 906
723 GGAAGTAAAGATTTTCTTGGGACTGATGATGCTCTCATATGATGACGAAAGAAAG 782
907 GGAAGCAAGAAATTTCTTCCGACAGATGATGCTCTCATATGATGACGAAAGAAAG 966
783 TGCTTCTGAGAGTGTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 842
967 TCATCTCGGAGATGCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1026
843 GTGTTGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 902
1027 GTCTTGTATGAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1086
903 GATTTTATGAGGCTTCTCTGAGAACTCAAGATTAAGTTGATGATGATGATGATGATGATGAT 962
1087 GATTTTATGAGGCTTCTCTGAGAACTCAAGATTAAGTTGATGATGATGATGATGATGATGAT 1146
963 GTACCGAATCTTTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1022
1147 GTTCCAGACGTTTCAACTTCCATTCGAGAGATGATGATGATGATGATGATGATGATGATGAT 1206
QY 1023 CTCGACTGTTGCG 1037
DB 1207 CTTCATGCGAACCG 1221
RESULT 13
CNS0A385 1294 bp mRNA linear HTC 06-FEB-2004
DEFINITION Arabidopsis thaliana Full-length cDNA Complete sequence from clone
GSLTRH932B09 of flowers and buds of strain col-0 of Arabidopsis
thaliana (thale cress).
ACCESSION BX827082
VERSION BX827082.1 GI:42461050
KEYWORDS HTC; GSLT cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
1 (bases 1 to 1294)
REFERENCE Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C.,
Menard,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,
Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.
Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:
A Combined Approach to Evaluate and Improve Arabidopsis Genome
Annotation

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1294)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)

COMMENT

The sequences are based on single pass reads.
 Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
 Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jallion O., Wincker P., Menard M., Crnaud C., Schachter V., Weissenbach J., Salanoubat M.
 URV INRA : Clepet C., Caboche M.
 Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences). 5 prime and 3 prime are assembled with Piprap.
 http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
 Length
 http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
 Location/Qualifiers

FEATURES
 source
 1..1294
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Col-0"
 /db_xref="taxon:3702"
 /clone="GSLTFRB93B09"
 /tissue_type="Flowers and buds"
 /plasmid="pCMVSPORT_6"
 1..1294
 /gene="At4g22930"

ORIGIN

Query Match 45.9%; Score 583.8; DB 3; Length 1294;
 Best Local Similarity 72.8%; Pred. No. 4,2e-150;
 Matches 753; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

3 GCAAAATGAGCTCTCAATCAACGAACTGATGATTTGATCTCCGTGATGAT 62
 200 GTAGAGTGGAGCTCAATCACTCAACGATGATTTGGCATCTTCCTGACGGT 259
 63 GATGTTCTTAAAGGAGTGTCTCTCAGAGTCAATCACTACCTTTGGGCGCAATGTCAG 122
 260 GATCTTCTTCAATGCTGTGTTCCCAACAGTGCAGATTTTGAAGACGATTTGATG 319
 123 CCAATTTGAAGCTCTTATCACTACGATGCTGCTGTGATACATACCGGAGGCGATA 182
 320 CCGAATCTGAAGCTCTCTGACCTTCACTGACGCTGCAATTAATTAACCGAAATTTATC 379
 183 TTGAATCTTACCTGTGATGATGATTTCAACCTCTTATGACCTTTATTTGACAGAT 242
 380 ATGAAGCTTTGCACTGATGATGATTTGATCTTATGACCTTTATTTGACTGAC 439
 243 ACAACCACTCTTATGAAATCAAACTAGCAAGAGAGCAGATGATTTGGGATGAAG 302
 440 AAAAATCTTAACTGAAAGATCAGGCTTGAAGGAAAGTGATGTTATGCGGTGAAG 499
 303 TTGTACCTGCTGCTGCAAGCAAAATTTCTCAAGATGATGATGATTTTGGGAG 362
 500 CTGTACCTGCGGAGCAACCAACTCTCAAGATGATGATGATTTTGGGAG 559
 363 TGTTTACCACTTCTACAGAAATGTTGATGATGATGATTTTGGGAG 619
 560 TGTTTACCACTTCTACAGAAATGTTGATGATGATGATTTTGGGAG 619
 423 GTTACTAATCTGAGGTTGACATGTTTATGATGATGATTTTGGGAG 482
 620 GTTACTAATCTGAGGTTGACATGTTTATGATGATGATTTTGGGAG 679
 483 AGACCTGTTGTCAGAAATTTTCAATTTGAAGTGTGATGATGATTTTGGGAG 542
 680 CAGCTCTAATCTCAACGCTTCCGAGCTGAAAGTATGATGATGATTTTGGGAG 739

QY 543 GATGCTTTAAGTTTGTGATTAATCTTGACATGAGATTTTGTGACGAACATGTCACCCCA 602
 DB 740 GATGCTTTAAGTTTGTGATTAATCTTGACGAAGAGGATGTTGGTGCAGAAAGTCAACCA 799
 QY 603 CAACATCTTGTGTTTGAACAGAAATCTCTTCAAGAGGAGGCTTCAACCGCATTAATAC 662
 DB 800 CAACATCTCTTCTCAACAGAAACGCTTTTCAAGAGGATTTCAACCTCACACTAC 859
 QY 663 TGCCCTCCAGTCTCTCAAAAGAGATCCACAGGAGGCACTTGTGATGATGAT 722
 DB 860 TGCTTTCCGTTCTCAAAAGAGAAATACATAGAGAGCATTTGTCAAAGCTGTAATAGT 919
 QY 723 GGAATGAAAGATTTTCTTGGAGATATGATGATGATGATGATGATGATGATGATGAT 782
 DB 920 GGAAGCAAGATTTCTTCTGAGCAATATGATGATGATGATGATGATGATGATGATGAT 979
 QY 783 TGTTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 842
 DB 980 ATATCTCGGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1039
 QY 843 GTGTTGAAAGAGAAATGCACTGCAAGAGCTTGAAGATTCATGATGATGATGATGATGAT 902
 DB 1040 GTCTTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1099
 QY 903 GATTTTATGAGGCTTCTTGAAGCAACTCAAAAGATTAAGTGAAGAGGAGGAGGAGGAG 962
 DB 1100 GATTTCTATGAGGCTTCTTGAAGCAACTCAAAAGATTAAGTGAAGAGGAGGAGGAGGAG 1159
 QY 963 GTACCCGATCTCTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1022
 DB 1160 GTTCCAGAGCTTTTCAACTTCCATTTGAGAGATGATGATGATGATGATGATGATGATGAT 1219
 QY 1023 CTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1037
 DB 1220 CTTCAATGCAACCG 1234

RESULT 14
 CDS0A2XU 1297 bp mRNA linear HTC 06-FEB-2004
 LOCUS Arabidopsis thaliana Full-length cDNA Complete sequence from clone
 DEFINITION GSLTFRB93B09 of Hormone Treated Callus of strain col-0 of
 Arabidopsis thaliana (thale cress).
 ACCESSION BX828087
 VERSION BX828087.1 GI:42460698
 KEYWORDS HTC; GSLT cDNA.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 1297)
 Castelli V., Aury J.M., Jallion O., Wincker P., Clepet C.,
 Menard M., Crnaud C., Quetier F., Scarpelli C., Schachter V.,
 Temple G., Caboche M., Weissenbach J. and Salanoubat M.
 Whole Genome Sequence Comparisons and 'Full-Length' cDNA Sequences:
 A Combined Approach to Evaluate and Improve Arabidopsis Genome
 Annotation
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1297)
 TITLE Direct Submission
 JOURNAL Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)

COMMENT
 The sequences are based on single pass reads.
 Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G.
 Genoscope members carried out sequencing and annotation : Castelli V., Aury J.M., Jallion O., Wincker P., Menard M., Crnaud C., Schachter V., Weissenbach J., Salanoubat M.
 URV INRA : Clepet C., Caboche M.
 Annotation is based on the June 2003 version of the Arabidopsis

genome released by MIPS (Munich Information Center for Protein Sequences). 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
length
http://www.genoscope.cns.fr/cgi-bin/g9b/g9b?source=Arabidopsis.

FEATURES

source

1. 1297
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Col-0"
/db_xref="taxon:3702"
/clone="GSLTPH5.02G03"
/issue_type="Hormone Treated Callus"
/plasmid="pCMVSPORT_6"
1. 1297
/gene="At4g22930"

gene

ORIGIN

Query Match 45.2%; Score 574.4; DB 3; Length 1297;
Best Local Similarity 72.8%; Pred. No. 1.7e-147;
Matches 754; Conservative 0; Mismatches 281; Indels 1; Gaps 1;

3 GCAAAATGAGCTCTCAATCAACAACCGATGATTGGCATCTTCATCCGATGGT 62
172 GTAGAGATGAGCTCAACATCACTCAACCTGATGATTGGCATCTTCATCCGATGGT 231
63 GATGTTCTTAAAGGAGCTGCTCTCAAGTCAATCACTTGGAGGGAATAGTCATG 122
232 GATCTTCTTCACTGCTGTTGTTCCCAAGTCCAGATATTTTAAGAGAGGATTTGATG 291
123 CCAAAATTTGAGGCTCTCTATCACTACCACTGCTGCTGTAGCATACCGAGGCGATA 182
292 CCGAATCTGAAGCTCTCTGACCTCACTCACTGACGCTCCATTTATCCGAAATTTATC 351
183 TTGAAATCTTACTCTGTTAGTGAATTCACCCCTTTTGAACCTTATTTAGACAT 242
352 ATGAAGCTTTGCTGATGAGAGTGTGATTCATCTTAAGACATTTATTTGACTGAC 411
243 ACAACGACCTCTATGAAATCAAACTAGCAAGAGAGAGCAGCTGTATTTGGGGTAG 302
412 AAAAATCTACTGAAGAGATCAAGCTTGCAAGGAAAGTGTGTTATGGGGTAG 471
303 TTGTACCTGCTGCTGTCACGACCAAAATTTCTAGATGAGTACTGATCTTTTGGGAG 362
472 CTGTAACCTGCGGAGCAACAACACTCTCAAGATGTGTCAAGGATCTTTTGGAAA 531
363 TGTTTACCACTTCAAGAAATGTTGAGATTAATGCTCTGCTGTTTCAATGAGAG 422
532 TGCTTACCGGTACTGAAAGATGTGTCAAAACAATGCTTTATCTGTTCAATGGGAG 591
423 GTTACTACTCTGAGTTGACATGTTGATAGAAAAGATTCATTTGAACCGTTCTA 482
592 GTCAAGATCCGATATCGATGTCTTTGACCGCAGAAAATCTTCAATTTGAGACGTTCTG 651
483 AGACCGTTGTGCAAAATTTCAAAATGAAGTGTGATGAGCAATTTTCAACAAT 542
652 CAGCTCTTAATCAACGCTTCCGAGCTGAAAATGATGAGAACATCAACATCATG 711
543 GATGCTGTTAAGTTGTTGAAATCTTGCACTGAAGATTTGTTGACGAACCTGTACCCCA 602
712 GATGCTGTTAAGTTGTTGAAATCTTGCACTGAAGATTTGTTGACGAACCTGTACCCCA 771
603 CAACATCTGTTTGAAGAAATTTCTCTTCAAGGGGCTTACCAACCGCATATATAC 662
772 CAACATCTCTTCTCAACGAGAGAGCGCTTTTCAAGGTGATTTCAACCTCAACAATAC 831
663 TGCCTTCAGTCTCAAAAGAGATCCACAGGAGGACCTTGTGTCACTGTAAAGAT 722
832 TGTCTTCCGTTCTCAAAAGAGAAATCATAGAGAGCAATTTGTTCAAAAGCTGTAACTAGT 891
723 GGAAGTAAAGATTTTCTTGGGACATGATGCTCTCATGATGACGAAGAAAGAG 782
892 GGAAGCAAGAGTTCTTCTCGGAGCAATAGTGTCTCAATGAAGAGATGAAAGAA 951

783 TGTCTTGTGATGATGCTGTTATTTTCAATGACCTGTAGCTTG-TCAGTATAGCGAA 841
952 TATCTCTGGATGATGCTGTTATTTTCAAGGCTCTCCATTCCTGCTCCCTATACGCCAA 1011
842 GGTGTTGAAAAGAAATGACATCGACAACTTGAAGATTCATGATTCATATGAGCC 901
1012 GGTCTTTGATGAGGGGGGCTGCGACAAAGTTGAAAGCTTTCACCAATTCATGAGCC 1071
902 AATTTTATGAGGCTTCTTGAAGAACATCAAGATTAAGTGAAGTAAACCCCATGAA 961
1072 TATTTCTATGAGCTTCCGAGAAACCTGTCAAGATCACTGAAGAAATCTCTTGAA 1131
962 GGTACCCGAAATCTTTTCTTATGATCAGAGATATATATCCATGTTGCTGTAAT 1021
1132 GGTTCAGAGCTTTTCACTTCCATTCGAGAGATGCTCTATGTTGCTGAGAAAC 1191
1022 GCTGACCTGTTGCCG 1037
1192 CTTCAATGCAACCG 1207

RESULT 15
CK251764 792 bp mRNA linear EST 30-JUL-2004
LOCUS
DEFINITION
EST735401 potato callus cDNA library, normalized and full-length
Solanum tuberosum cDNA clone POCB96 5' end, mRNA sequence.

ACCESSION
CK251764
VERSION
CK251764.1 GI:39805104
KEYWORDS
EST.

SOURCE
ORGANISM
Solanum tuberosum (potato)

Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamiales; Solanales; Solanaceae; Solanum.

1 (bases 1 to 792)

Buell, C.R., Hart, A., Zismann, V., Karanicheva, S.A. and Baker, B.
Generation of ESTs from potato callus tissue

Unpublished (2003)

CONTACT: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA

Email: potatobu@igrr.org

Clones can be requested from the University of Arizona Genomics
Institute via http://genome.arizona.edu/orders/

Seq primer: ATT TAG GTG ACA CTA TAG.

Location/Qualifiers

FEATURES

source

1. 792
/organism="Solanum tuberosum"
/mol_type="mRNA"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="POCB96"
/issue_type="callus"
/lab_host="DH10B-TonA"
/clone_id="potato callus cDNA library, normalized and
full-length"
/note="Vector: pCMVSPORT6.1; Site 1: EcoRI; Site 2: NotI;
supplier: RNA was isolated from Solanum tuberosum var.
Kennebec callus tissue grown on solid media."

ORIGIN

Query Match 44.1%; Score 561; DB 7; Length 792;
Best Local Similarity 100.0%; Pred. No. 7.5e-144;
Matches 561; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2 TGCAAAATGAGCTCTCAATCAACAACCGATGATTGGCATCTTCATCCGATGG 61
232 TGCAAAATGAGCTCTCAATCAACAACCGATGATTGGCATCTTCATCCGATGG 291
62 TGATGTTCTTAAGGAGATTTGTTCTTCAAGTCAATCACTTTTGGAGGGAATAGTCAT 121
292 TGATGTTCTTAAGGAGATTTGTTCTTCAAGTCAATCACTTTTGGAGGGAATAGTCAT 351

